MODELLING APPEARANCE AND GEOMETRY FROM IMAGES

A THESIS SUBMITTED TO THE UNIVERSITY OF MANCHESTER
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE FACULTY OF ENGINEERING AND PHYSICAL SCIENCES

2011

By
Francisco Melendez
School of Computer Science

Contents

Words: 56.109

List of Tables

List of Figures

10

12

Abstract

Acquisition of realistic and relightable 3D models of large outdoor structures, such as buildings, requires the modelling of detailed geometry and visual appearance. Recovering these material characteristics can be very time consuming and needs specially dedicated equipment. Alternatively, surface detail can be conveyed by textures recovered from images, whose appearance is only valid under the originally photographed viewing and lighting conditions. Methods to easily capture locally detailed geometry, such as cracks in stone walls, and visual appearance require control of lighting conditions, which are usually restricted to small portions of surfaces captured at close range.

This thesis investigates the acquisition of high-quality models from images, using simple photographic equipment and modest user intervention. The main focus of this investigation is on approximating detailed local depth information and visual appearance, obtained using a new image-based approach, and combining this with gross-scale 3D geometry. This is achieved by capturing these surface characteristics in small accessible regions and transferring them to the complete façade. This approach yields high-quality models, imparting the illusion of measured reflectance.

In this thesis, we first present two novel algorithms for surface detail and visual appearance transfer, where these material properties are captured for small exemplars, using an image-based technique. Second, we develop an interactive solution to solve the problems of performing the transfer over both a large change in scale and to the different materials contained in a complete façade. Aiming to completely automate this process, a novel algorithm to differentiate between materials in the façade and associate them with the correct exemplars is introduced with promising results. Third, we present a new method for texture reconstruction from multiple images that optimises texture quality, by choosing the best view for every point and minimising seams. Material properties are transferred from the exemplars to the texture map, approximating reflectance and meso-structure. The combination of these techniques results in a complete working system capable of producing realistic relightable models of full building façades, containing high-resolution geometry and plausible visual appearance.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institution of learning.

Copyright

i. The author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the "Copyright") and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii. Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with the Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from time to time. This page must form part of any such copies made.

iii. The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the "Intellectual Property") and any reproductions of copyright works in the thesis, for example graphs and tables ("Reproductions"), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv. Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://www.campus.manchester.ac.uk/medialibrary/policies/ intellectual-property.pdf), in any relevant Thesis restriction declarations deposited in the University Library, The University Library's regulations (see http://www.manchester-.ac.uk/ library/ aboutus/ regulations) and in The University's policy on presentation of Theses

Acknowledgements

First I want to thank Professor Roger Hubbold for his supervision, whose experience, patience to correct my spelling mistakes, and advice during the last four years made this thesis possible and are enormously appreciated. I also want to express my gratitude to Dr. Mashhuda Glencross, my unofficial co-supervisor, for her invaluable encouragement, her trust, and her friendship. Thanks to Mr. Nick Glencross, for his support in the last few months, and Mr. Greg Ward, who provided much help with data capture and code. I also would like to thank Mr. Toby Howard for being my thesis advisor. Special thanks to Dr. Jun Liu and Omar Alvi for the always insightful discussions about life and research. I wish to mention in this acknowledgements Diego Gutierrez, somehow responsible for me coming to the U.K. to study a Ph.D. To friends and colleagues in the School of Computer Science: Shaobo Hou, Rezwan Sayeed, Farzaneh Sarafraz, Salil Deena, and Timo Kunkel.

Quiero recordar también aquí a mis amigos en Manchester que me han aportado tantos buenos momentos y tantas experiencias: Ana Chalas, Raquel, Esther, Michelle, Brian, Thodoris, Nuno, Francisco Martin, Gloria, Diego, Clara, Silvia, Susana, Ivan, Graciano, Catarina, Luis, Rebecca, Roi, Laura Marino, Edu y Bea, Lucia, Alison, Eleni, Amanda, Francisca, Lea, Sahar, Lukasz, Uma, Elena Davitti, Sandra, Leonie, Matt, Rocio, Gisela y Roberto, Beatriz Hernández de Madrid, Elena Martín, Ana Hidalgo, Toni, Lucia y Cesar, Philip ... y Daria.

También agradecer su apoyo a mis amigos en Zaragoza, Aitor, Alberto, Oscar, Jorge y Carlos.

Por último a mi querida familia: mis abuelas Ubalda y Edelia, mis padres Olga y Antonio, mis hermanos Irene y Antonio, y en el recuerdo mi abuelo Adolfo.

Introduction

ACQUISITION *of realistic and relightable 3D models of large outdoor objects, such as buildings, requires the modelling of detailed geometry and visual appearance. Recovering these is a complex task, needs specially dedicated equipment, and can be very time consuming. This thesis investigates how to recover high-quality models from images, using simple photographic equipment and modest user intervention. It explores how to enhance such models with surface detail to provide a compelling visual appearance. In this introductory chapter, we motivate and focus our approach, outline the contributions of this thesis, and present its structure.*

1.1 Digitising the Real World

Creating digital 3D models of existing objects and environments has become an important aspect of producing realistic computer graphics. Digitising buildings is of importance to a range of possible applications including architectural visualisation and planning, computer games, film post-production, archaeology and cultural heritage. The recovered 3D models can be augmented with CAD models of new buildings, or alterations to existing structures, such that the visual impact of the new designs can be assessed. Games and films demand models of increasing levels of realism as well as reducing the cost and effort dedicated to generating them. For buildings of historic significance, analysis of the visual impact of changes, such as the construction of a new visitors centre, or other extensions, is of primary importance in the planning process.

These applications need to be capable of rendering the synthetic scenes under changing lighting as well as view point, – for example, a historic building may require simulation under floodlighting – necessitating digitising not only the geometry of the objects, but also their visual appearance: the way surfaces reflect light at each

point. Existing techniques, for building complete 3D models, involve a trade off between the quality of the recovered models and the amount of effort and cost to create them. The accepted way to create a high-quality model today, is to use a laser scanner and reflectance measurement equipment, but this requires both expensive instruments and significant effort for both data capture and final model assembly. In contrast, image-based approaches offer the promise of low cost equipment, simple capture requirements and automated model assembly. The challenge for image-based methods is to build complete models with sufficient local detail that compare favourably with more labour intensive approaches.

Acquisition of detailed 3D geometry from photographs has proven to be a complex problem. Existing flexible and automatic systems aimed to capture the global structure – the *gross-scale geometry* – of building façades from a set of views. However, few current methods exist to easily capture and incorporate local detailed surface –*meso-scale geometry* or *meso-structure* – such as cracks and protrusions in stone walls, necessary to create realistic appearance under novel lighting. Furthermore, these methods, such as photometric stereo, require high-resolution images, normally captured at a close range, restricting their usability to small portions of the surface. Mostly, surface detail is conveyed by textures recovered from images, *whose appearance is only valid under the originally photographed viewing and lighting conditions.*

Visual appearance can be even more difficult to digitise for large objects under outdoor lighting conditions. Reflectance properties can be complicated to measure and express for even a single point, and typically vary across the surface of an object. These can be estimated from digital photographs of the objects captured from different viewing and illumination directions. Unfortunately, large outdoor scenes make it complicated to control the illumination and viewpoints. During daytime, the lighting conditions change continuously, and large surfaces are most easily photographed from ground level making it impractical to cover a full range of angles. An important aspect of visual appearance is its *albedo*, i.e. the surface colour under a full spectrum white light. Albedo is a simplification of the complete reflectance function of a surface, but acquiring it involves factoring out reflectance and lighting from the observed photographs. Although texture includes colour information, separating albedo from lighting is an inherently ill-posed problem. This separation is key to conveying visual realism of textured surfaces under novel lighting.

1.2 Geometric Detail and Visual Appearance from Photographs

In this thesis we address the problem of constructing visually faithful models of building façades from photographs. Our objective is for reconstructed models to appear as much like the real scenes as possible under varying viewing and lighting conditions, such as on a cloudy day, in bright sunshine, or at night-time when floodlit. The main focus in this investigation is on approximating detailed surface meso-structure (local depth information) and visual appearance (albedo), and combining this with gross-scale 3D geometry obtained using a different image-based approach. This is achieved by capturing these surface characteristics in small accessible regions, and then transferring them to the complete façade.

We envisage a scenario where a user takes several pictures of a building from different viewpoints, and sample images of materials at a close range, comprising a flash and a no-flash photograph of each material. Together, the flash and no-flash pair provide exemplars for each type of texture in the building. Global scene structure and camera parameters are recovered automatically using a previously developed system, extended here to handle larger models. The material exemplars are also processed automatically, using image processing algorithms, to recover the meso-structure and reflectance of the textures. Then, using the novel techniques developed in this thesis, the user transfers the surface detail and appearance to the complete model. Such a transfer process poses difficulties in reconstructing an appropriate texture map for the complete building, identifying different materials in the façade, associating them with the corresponding exemplar, and finally, effectively transferring the desired material characteristics. This approach yields high-quality models, imparting the illusion of measured reflectance. In Figure 1.1, our reconstructed model of a historical English castle in the city York – Clifford's Tower – is rendered under novel lighting conditions and compared with a photograph, taken from the Internet [1].

Thus, we present a complete image-based system that facilitates recovery of gross-scale geometry, local surface structure, and surface albedo, to create highly detailed 3D models of building façades from photographs. A key aspect of our process is that it uses only simple digital SLR equipment and has low data capture and labour requirements. Investigations to further automate the process are also described.

---

[1]Source: http://tigg-stock.deviantart.com/art/Castle-stock-10-73841672

Figure 1.1: *Clifford's Tower, York. The image on the right is a photograph taken from the Internet. The image on the left is a model reconstructed using our approach, rendered under novel lighting conditions, and seen from a similar view point to the photograph.*

1.3 Summary of Contributions

To the knowledge of the author, this is the first attempt to apply exemplar-based texture transfer techniques to the problem of recovering meso-scale geometry and visual appearance for outdoor scenes. This is considered the main contribution of this thesis. Other novel and original work presented in this thesis includes:

- **Two novel algorithms for meso-structure and albedo transfer**, where these material properties are captured for small exemplars, using an image-based technique. These methods are studied analytically and tested.

- **An extension of the transfer to several materials and larger surfaces**. We present an interactive solution to solve the problems of performing the transfer over a large change of scale and to the different materials contained in a full façade. Aiming to completely automate this process, a novel algorithm to differentiate materials in the façade and associate them with the correct exemplars, is introduced with promising results.

- **A new method for texture reconstruction from wide-baseline multi-view images**, based on graph optimisation. We formulated the problem as a 'picture mosaicking' to reduce artifacts and maximise the quality, producing high-resolution textures. Material properties are transferred from the exemplars to the texture map, approximating reflectance and meso-structure.

- **A complete working system** capable of producing realistic models of full building façades, containing high-resolution geometry and plausible approximations of

visual appearance. The system developed uses easy to capture photographic data, and is largely automatic requiring only modest user interaction for some parts of the process. This extends the previous gross-scale geometry reconstruction system, which could only capture untextured low/middle resolution models.

1.4 Thesis Outline

In this thesis we address the difficulties of capturing and estimating geometric detail and appearance for building façades from images. Chapter 2 introduces and defines the specific problems of reconstructing outdoor scenes. We begin by commenting on the digitisation of geometry, separating global geometry and surface detail. We review, for context, the techniques for gross-scale geometry reconstruction from images, and introduce the multi-view modelling system that we used and extended during the completion of this thesis. Then, we present meso-structure capture techniques and the limitation of the previously published techniques. Capturing meso-scale geometry overlaps with the acquisition of visual appearance, which is also discussed in this chapter with particular emphasis on the problems of outdoor conditions. This serves as justification to investigate our novel transfer approach, which is presented together with an overview of the complete reconstruction system.

Chapter 3 reviews the Depth Hallucination technique [2] for capturing material exemplars of textured surfaces. After analysing this process, and the characteristics of the data, we introduce two different methods to transfer the captured surface detail and visual appearance: Histogram Matching and Transfer by Analogy. This process is performed and evaluated on a one material-to-one material basis, and for small samples of texture. The extension to transfer several materials is presented in Chapter 4. We discuss our practical solution to perform the transfer of multiple materials contained in a façade, using semi-automatic techniques. Targeting the complete automation of this process, we also introduce an extension to the Transfer by Analogy method to associate exemplars and materials in the façade automatically.

Once the transfer techniques have been presented, we address their application to texture maps of complete buildings in Chapter 5. We discuss the requirements of an optimal texture map and present a novel approach for its acquisition. Then, we analyse the effects of the deformations caused by the texturing process in the application of our

---

[2]Although the author collaborated in this work published in [GWJ$^+$08], the Depth Hallucination technique is not part of the contribution of this thesis

transfer techniques. The end of Chapter 5 is dedicated to the combination of recovered surface detail with the gross-scale geometry reconstructed previously.

An overall evaluation of the reconstruction system is carried out in Chapter 6. We show the resulting reconstructions of three historic buildings that contain the typical challenges of outdoor scene reconstruction. We compare our models, enhanced with the meso-scale geometry and visual appearance recovered with our transfer techniques, with the models recovered other approaches and with photographic images of the buildings, illustrating the effectiveness of our approach through visually realistic reconstructions. Finally, Chapter 7 summarises the thesis and suggests future lines of investigation.

Reconstruction of Outdoor Building Scenes

THIS *chapter introduces the context of 3D reconstruction and de-fines the research problem, starting point, pipeline and moti-vation of the novel techniques developed in this thesis. Furthermore, it reviews previous research in geometry and reflectance acquisition and justifies the necessity of a novel approach for our specific re-quirements.*

Representation of a scene in computer graphics consists of three main components: geometry, visual appearance, and illumination. To create renderings of existing objects under changing lighting as well as view point, it is necessary to digitise their geometry and visual appearance. Normally, and in the context of this thesis, geometry refers only to surface geometry of opaque objects, and visual appearance is limited to surface reflectance properties. Modelling and digitising translucent objects and materials, and rendering physically complex phenomena such as fluorescence, phosphorescence, subsurface scattering or polarisation, require specialised techniques that are out of the scope of this thesis and of most rendering applications.

Our focus is on techniques to recover digital models of large structures, such as buildings, under outdoor conditions using only photographic data. Large scenes and the use of photographs as input data lead to a series of specific difficulties for recovering both geometry and reflectance properties. Detailed large-scale objects require a large number of images to be modelled appropriately, so resolution and hardware requirements quickly become an issue. On the other hand, recovering reflectance from photographs requires knowledge or control over the lighting, which is challenging under natural lighting conditions. In fact, natural illumination conditions are considered uncontrollable, and to date, measuring them requires special equipment.

This chapter introduces the problems and solutions for acquiring relightable models of existing outdoor scenes. We present an overview of current approaches for digitising

geometry and appearance which motivates our novel approach for easy surface detail and reflectance recovery based on texture transfer methods.

2.1   Digitising Geometry

Geometry in computer graphics refers to a 3D representation of object surfaces. The usual representation is a mesh which contains a set of points in 3D space (vertices) and combinations of them defining planar faces (facets). The number of vertices and facets of this piecewise representation of the geometry will define its resolution and level of detail.

We make a distinction between two kinds of geometry depending on the level of detail. In the example of a building, we define *gross-scale geometry* as the global 3D structure, such as walls and roofs, meanwhile the fine detail, like bricks, cracks, and other texture patterns, is denoted *meso-scale geometry* or *meso-structure*. Meso-structure assumes an underlying planar surface and can be captured using different techniques to those used to recover gross-scale geometry. Modelling this kind of geometry is important to reproduce self-shadowing effects in textured surfaces, which provide visual realism and a more accurate appearance.

A common approach to make geometric models tractable in terms of memory is to have a middle-low resolution geometric meshed model of the gross-scale geometry, and represent the high resolution detail or meso-scale geometry as a texture map. Texture maps are images that code superficial information, such as per pixel surface orientation, relative displacement with respect to the gross-scale geometry, and colour.

The following sections review separately current techniques to recover gross- and meso-scale geometry.

2.1.1   Gross-scale Geometry

There exist two main methods for reconstructing geometric models of existing objects: *Range-based modelling* and *Image-based modelling*. Range-based modelling uses active sensors (typically laser scanners) that can provide highly accurate depth measurements in the form of 3D point-clouds which can be meshed to a final 3D model. Scanners work at different ranges and resolutions, and large objects require several scans which need to be aligned and filtered. The technology is mature and it is considered

the most reliable way to acquire geometric models, having become popular for applications such as cultural heritage [RBMT98, IS01, GGV+04, DTG+04b, IOT+07]. However, scanners need careful calibration, require expert knowledge of the capabilities of each technology, and their high cost inhibits their widespread adoption, especially for individual users.

Image-based modelling (IBM), on the other hand, offers the possibility of low cost solutions and has attracted researchers and industry interest in the last two decades. IBM is the process of creating three-dimensional models from a collection of images or video. Computer vision algorithms, with different grades of user interaction, estimate camera positions, match features between images and extract 3D structure from the multiple views. Impressive results can be produced with interactive systems [DTM96, EH02, GHCH03, DTC04, SSS+08a, XFT+08] where the user can model scenes using the images and some automatically recovered 3D information as guides. These models are normally composed of simple geometric primitives such as planes, spheres, and cylinders. More complex structures require skilled users and some time to model. Advances in the theory of multi-view geometry gave way to several automatic image-based modelling systems from uncalibrated photographs including those of [FZ98, Nis05, PvV+04, RP05, LQ05]. Further research was focused on making the relative baseline between images wider still recovering quasi-dense point-clouds [BL05, Liu07]. By using similar techniques, recent work took advantage of large photo collections from the Internet to reconstruct buildings [GSC+07, SSS06, SSS08b] and even cities [ASS+09]. However, those approaches only recover sparse 3D representations of the scenes. Furukawa et al. [FCSS10] subsequently applied dense matching to similar large data collections, improving the density of the final point-clouds. These techniques are very flexible, low cost, and some of them, totally automatic. They provide good models of the high-level 3D structure. However, when trying to capture more detailed geometry, dense matched point clouds are noisy in comparison to laser-scanned data, and produce artifacts at the meso-structure level.

In summary, interactive image-based modelling techniques are very efficient at providing relatively simple low-resolution geometric models containing simple geometrical primitives, but do not provide enough detail to compute self-shadowing effects. Automatic techniques can provide quasi-dense point clouds in a similar fashion to laser scans, but these point clouds are not regularly distributed, and resolution and accuracy are limited. Consequently, when re-lighting the geometric models obtained in this way, little detail can be reliably computed; where such detail is attempted, it is often noisy,

and gives poor or unreliable self-shadowing effects. We conclude from this review that automatic, or quasi-automatic, Image-based modelling techniques provide good gross-scale geometric models, but dissatisfy when a higher-level of detail is demanded.

In the following section, we overview a state-of-the-art system for automatic scene reconstruction from wide-baseline photographs, previously developed in our group [Liu07]. This system was used to recover gross-scale geometry for the models presented in this thesis. We also present solutions to two limitations that arise when reconstructing models of complete buildings with this system: registration of partial models and filtering of unreliable 3D points.

2.1.2 Gross-scale Geometry from Wide-baseline Images

To create our models, we use a system previously developed in our group named *Helios*, described in detail by its author in his doctoral thesis [Liu07]. In principle, any other similar system can be used to recover the gross-scale geometry. Helios automatically detects features [Low04] in the input images (Figure 2.1), and matches these across the sequence using robust methods to estimate camera parameters [Nis04] and the 3D position of the matched points. By using the reconstructed feature points as seeds, it propagates the matching to the neighbouring pixels using *Oriented Normalised Cross Correlation* [Liu07]. After a bundle adjustment stage, where 3D point positions and camera parameters are optimised iteratively, the system provides a quasi-dense 3D reconstruction from a collection of still images in the form of a 3D point cloud, camera positions and orientations for every view, and correspondences between the reconstructed 3D points and the 2D pixels in the view where they are visible. Since the raw output (a 3D point cloud) from the reconstruction process is somewhat noisy and not uniform, we obtain a gross-scale model by estimating a normal vector per point, and then applying Poisson surface reconstruction [KBH06] to obtain a triangulated implicit surface enforcing continuity and smoothness.

This reconstruction system suffers from several limitations which are common to similar techniques running on regular desktop computers. First, scalability is constrained by the memory available in a desktop computer (4GB nowadays), which determines the number and resolution of the images that the system can handle in a single reconstruction. The second limitation is the noise in the reconstructed point cloud, which also depends on the number of views and image resolution, as well as texture features available in the building.

We propose solutions to these two limitations. First, we present a simple robust

Figure 2.1: *Sequence of multi-view photographs used for reconstruction and resulting point cloud.*

technique for aligning partial reconstructions. Second, we use the uncertainty model described in [Liu07] to reject unreliable propagated points.

Scalability to Full Models

The incremental nature of the reconstruction algorithm can, in principle, handle an arbitrary number of input images. In practice, this is limited by available memory. The bundle adjustment stage is the most memory-demanding component in the reconstruction pipeline, increasing linearly with respect to the total number of 3D points and number of cameras. The number of points that can be acquired in a single reconstruction, is approximately one million in a 4G RAM memory desktop. Although the current implementation re-samples the quasi-dense 3D points before bundle adjustment, it is necessary to have a method to deal with large scenes. We designed a method to merge partial reconstructions as suggested in [Liu07]: "merging partial reconstructions may be a viable option and needs further investigation".

Aligning and merging partial reconstructions is a common problem for scanned data [RL01, Fit03, BL04]. A widely used algorithm is Iterative Closest Point(ICP) [BM92] and variants [RL01, Zha92] which starts with an initial guess of the rigid transform between datasets, and iteratively refines the transform by generating pairs of points on the models and minimising the least squares error between the two point sets. Several
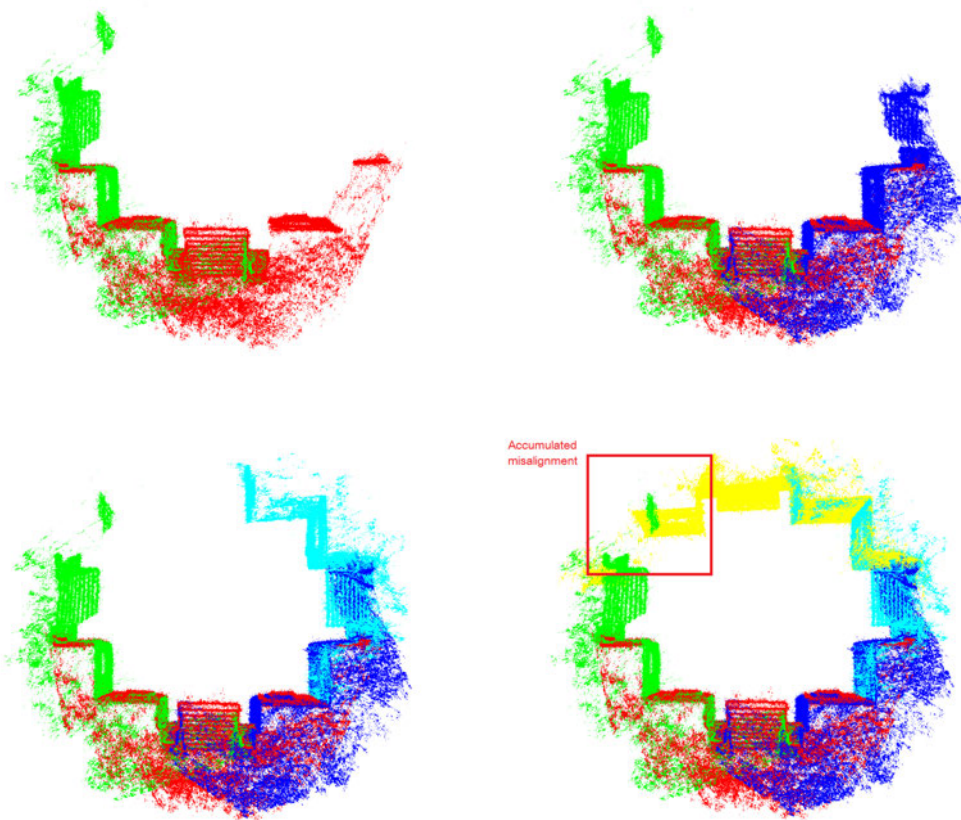
Figure 2.2: *Point-clouds combined into a single point cloud. Each colour represents a different sequence. We incrementally align new sequences. The last sequence presents some misalignment (highlighted by a red square) with the first sequence due to distortion of the input data. Note that first and last sequences are not aligned directly but by the intermediate point-clouds.*

algorithms have been proposed for simultaneously registering a collection of scans. Chen and Medioni [CM91] perform global alignment by incrementally aligning new range scans to all previous ones. This approach suffers from errors accumulating as successive scans are added. Other approaches including [KLMV05, IMNI08] align all scans simultaneously which distributes the errors globally. When no initial guess is available, features need to be detected and matched between point-clouds. This is generally a difficult part of registration techniques. We avoid this problem by introducing a common image in consecutive reconstructions. Then, we find the common reconstructed points by comparing the reprojected points in 2D from both sequences in the same image. Having the correspondences, we need to find the rigid transformation

between both models. This transformation is a 4×4 matrix that define the translation, rotation and scaling. We used the robust RANSAC(Random Sample Consensus) method [FB81] to get a good initial estimation and reject outliers. Then we use non-linear least-squares minimisation to find the transformation that best fits the data. This method deals appropriately with outliers and noisy data. As in the work of Chen and Medioni [CM91], our alignment algorithm is sequential and accumulates error, however the misalignment error is small compared with the distortion of the point-clouds due to errors in camera estimation. This method does not correct for such distortions. Figure 2.2 shows the progressive alignment of several partial sequences of the *Eagles and Jaguars* model. The red square in the bottom-right image, shows that the first and last sequences have some misalignment offset, however, the alignment in the rest of the overlapping areas presents minimal errors.

We tried solving the optimisation process globally, but the results presented larger distortions. Since the main source of misalignment is the distortions of the point clouds, a global optimisation distributes the error over all the partial models, producing large misalignments also in all the overlapping areas. Correcting for input data distortion could be solved with aligning methods that allow for non-rigid transformations [IGL03, BR07]. This is a more difficult problem since the distortions in the model are unknown a-priori. Allowing for non-rigid transformations would require, in this case, establishing constraints like symmetry, or user guidelines, so only the distorted models are corrected.

In summary, our algorithm sequentially aligns partial models dealing appropriately with noisy data and outliers. It improves the scalability of the system allowing it to reconstruct larger scenes. However it is limited by the correctness of the input data and further research is necessary in order to correct distorted models.

Filtering Noisy Points

Both merging partial models and the meshing algorithm, are robust methods that can deal with noisy data. However, removing noisy data improves their performance. Filtering noisy point data is a common problem for laser scanned data [SBS05, WPH$^+$04]. We propose a simple threshold-based filtering method, that uses the uncertainty computed from the Structure from Motion process, that can complement other filtering techniques.

Estimated depth for points matched across several views, that differ little in their baseline, cannot be accurately constrained, and the uncertainty about the position of

this point will be large. On the other hand, when the views have a large angular difference between them, the depth can be estimated more accurately. We use the uncertainty to filter the point cloud and reduce the amount of noisy data.

The uncertainty model is defined [Liu07] and provides a measurement of the confidence about every point in every direction in 3D. Uncertainty is modelled as an ellipsoid, representing the uncertainty in every direction. We take as the uncertainty value for every 3D point, its largest uncertain value, which coincides with the largest Eigen value of its covariance matrix. We apply a threshold-based filter to discard unreliable points.

(a) Point cloud seen from an input view point.

(b) Point cloud seen from a novel view point. Note the noisy points in the red circles.

(c) Filtered point cloud using 0.5 threshold value

(d) Filtered point cloud using 0.1 threshold value
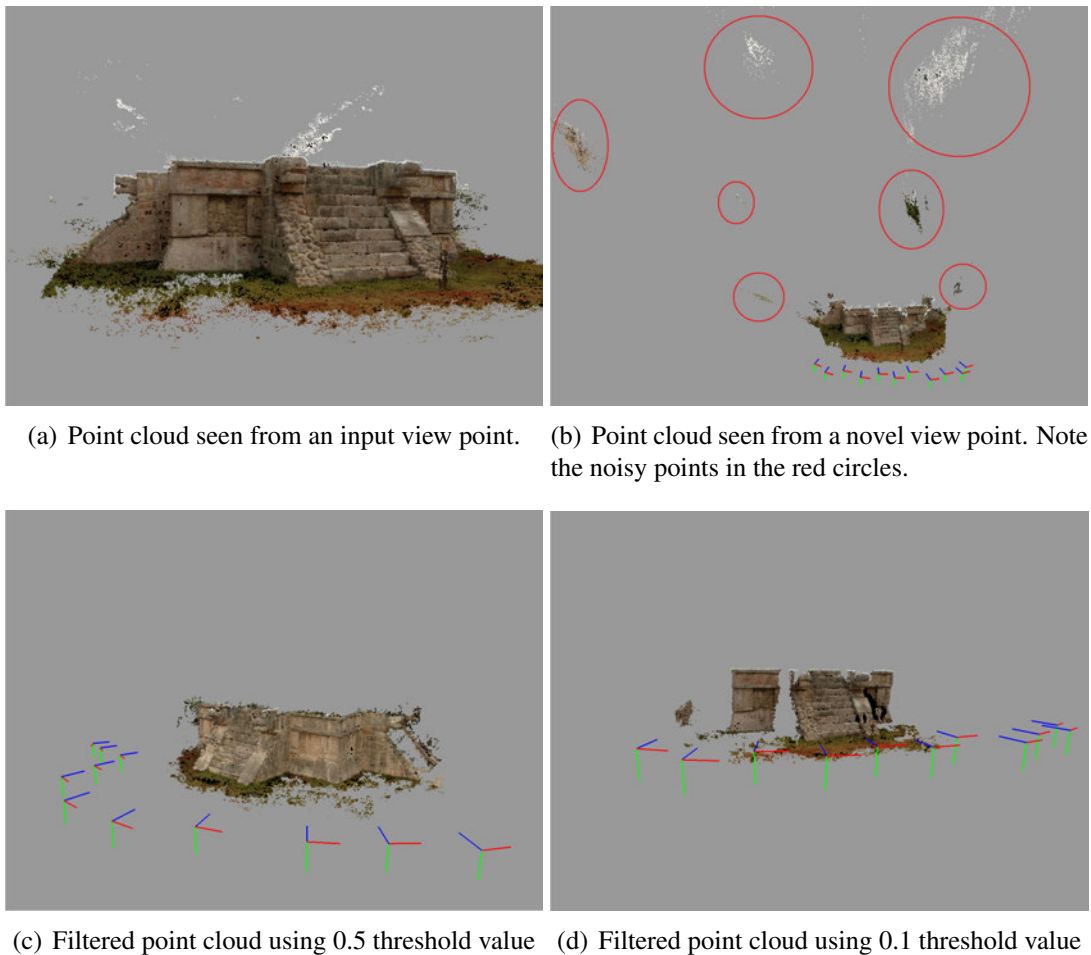
Figure 2.3: *Example of the cleaning effect on a reconstructed point cloud.*

Figure 2.3 shows the results of applying our filtering technique. Figures 2.3(a) and 2.3(b) show the model before filtering, where very distorted points can be observed in the highlighted circles. All these points are filtered out using uncertainty. If the

user wants only the best points, for instance if this data is available in other partial sequences, we can set a smaller value for the threshold. In figure 2.4 we show the distribution of the uncertainty data for the same partial model.
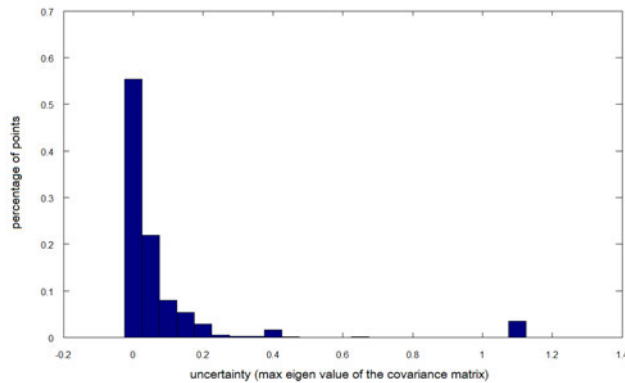
Figure 2.4: *Histogram of the uncertainty. For this model, around 90% of the points have an uncertainty less than 0.2. The values over 1 are cropped to 1.1 for visualisation purposes.*

The appropriate threshold is dependent on the quality of the reconstructed point cloud. We need to find a trade-off between the number of points remaining and their quality. Studying the values of our uncertainty measure in Figure 2.4, we see that most of the points have values between 0 and 0.5. Applying a threshold value of 0.5 (see Figure 2.3(b)) removes the highly distorted points keeping a well sampled model in all areas. A value of 0.1 (See Figure 2.3(a)) leaves holes in the model that would lead to artifacts in the final mesh.

We typically use a threshold value between 0.1 and 0.5, depending on the sequence. Values larger than 1 are normally useless. Sometimes, by keeping uncertain points in areas where the point cloud is very sparse, we obtain a better final mesh.

Meshed Models

After cleaning and merging the partial models from the image-based modelling system, we use Poisson surface reconstruction presented by Kazhdan et al. [KBH06] to create an implicit surface from the point cloud. Figure 2.5(b) shows the mesh obtained with this method. The algorithm tries to create a closed surface which sometimes produces extra geometry like the dome above the platform in the Figure. This excess geometry is then cleaned and smoothed by using open source software Meshlab [1]. The

---

[1]http://meshlab.sourceforge.net/

cleaning process is carried out manually, with the tools provided by the software. The cleaning process requires modest skills and effort, although in principle, it could be automated by comparing the mesh with the reconstructed point cloud. The final result, in Figure 2.5(b), constitutes our gross scale geometry.

(a)   (b)

(c)
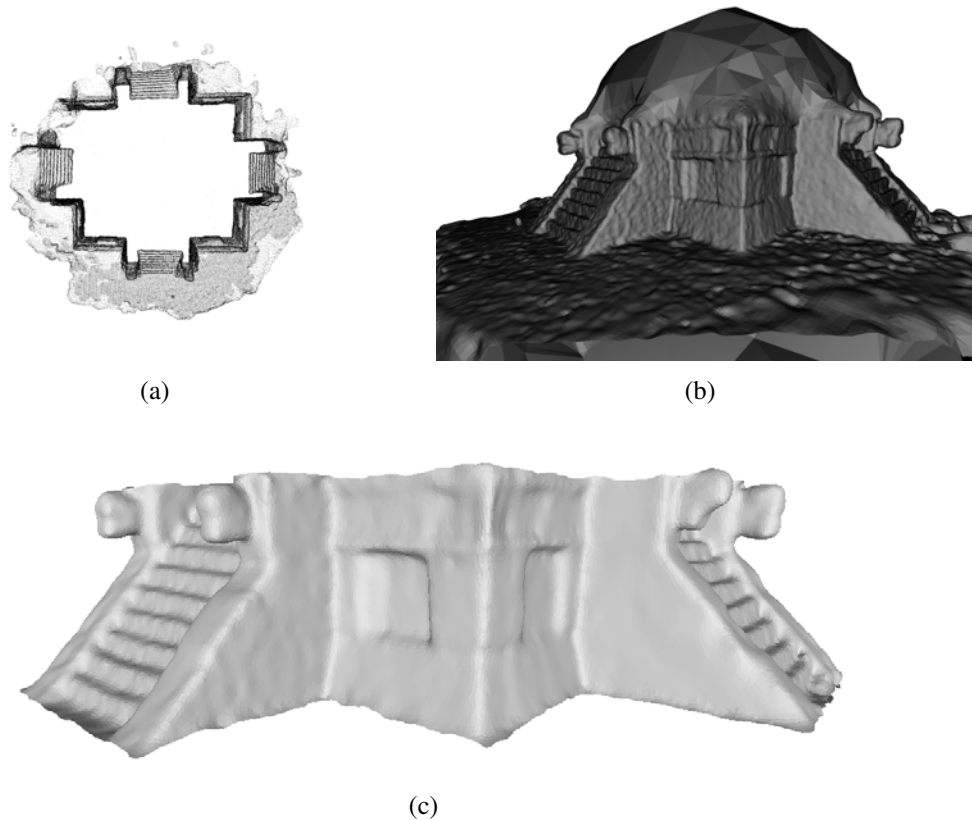
Figure 2.5: *(a) Vertexes of the implicit surface after meshing. (b) Reconstructed mesh resulting from the meshing algorithm. (c) Mesh after cleaning and smoothing.*

We observe that the gross scale geometry contains the global structure of the building (in this case a Mayan platform), but lacks surface detail or meso-scale geometry.

Meso-scale geometry could be captured using similar image-based methods such as multi-view stereo, by using close views, at the cost of capturing a vast number of pictures. However, since only the local detail is required, we can make safe assumptions about the global geometry, for instance that is comprised of largely planar surfaces, and use a different approach capture it. The next section reviews methods to recover *meso-scale geometry* at similar, or even higher resolutions, to laser-scanned models.

2.1.3 Meso-scale Geometry from Images

We define meso-scale geometry, meso-structure or textural geometry as the local variation of the relief in globally planar surfaces. Many models used in graphics applications do not explicitly model this type of detailed geometry and instead use simple colour texture maps to give the appearance of detail on the surfaces. Texture maps can produce very realistic results, but they do not hold when changing view point or changing lighting conditions. Effects like self-shadowing and correct appearance in oblique views, are important to convey realism in computer graphics, and require meso-structure to be modelled.

The techniques previously presented for acquiring gross-scale geometry, such as multi-view stereo and range measurements with laser scanners, can be applied to meso-scale geometry acquisition by using higher resolution images. However, the nature of the meso-scale geometry – small depth disparities over a planar surface – allow other image-based techniques to estimate geometric detail from a single view point, recovering depth per pixel without the holes that usually appear in multi-view stereo and laser scanning. In this section we review the ideas behind *photometric-stereo* and *Shape-from-shading*.

Photometric stereo [Woo80] recovers surface orientation from a single view by changing the lighting direction. Assuming a surface whose reflectance does not depend on the view point, i.e. reflects the same amount of light in every direction, and knowing the lighting directions with respect to the camera, three images are sufficient to recover surface normal direction and reflectance. Multiple extensions to the original algorithm have been proposed, allowing for general reflectance properties [Geo03], resolution at sub-pixel level [TLQ06], and relaxing the illumination conditions to general unknown lighting [BJK07]. These techniques require the same view point under several lighting conditions, so are mainly restricted to controlled environments. More flexible capture systems have been developed emphasising the easy capture and general application. Rushmeier and Bernardini [RB99] used a portable rig with five tungsen-halogen light sources, so they could ignore up to two lights that caused specular reflections or shadows. A simplification of this method presented by Ward and Glencross [WG09], which is an especially flexible capture system, uses a three flash rig. They used crossed polarisers to avoid specular highlights, and interpolated normals in shadowed areas. A similar system was presented by Paterson et al. [PCF05]. These researchers introduced a physical frame around the target which allowed them to take the pictures from different viewpoints, recovering full, complex reflectance. However the approach limits

the samples to a small portion of the surface (approximately $20 \times 30cm$) .

Shape-from-shading solutions aim to acquire 3D depth information from a single image [KvD83, Hor89, MM89, HF98a, HF98b, PF05]. This is an under-constrained problem, since, as we have seen, at least three images are necessary. Numerous shapes, surface reflectances, and lighting conditions can give rise to the same shading pattern [BKY99]. However, shape-from-shading approaches are attractive for recovering meso-scale geometry as they do not require special equipment or lengthy data-capture processes. These techniques infer surface orientation under a single known lighting configuration, normally a unique light source at infinity. They also assume a Lambertian surface of uniform albedo equal to 1. With further assumptions about the underlying shape, lighting direction can be estimated [ZC91]. A large body of literature on the topic of shape-from-shading exists, and we refer to published surveys for a review of existing methods [ZTCS99, DFS07].

A variation of the classical shape from shading model analyses the perception of shape under diffuse illumination, where the light arriving is the same for all directions. Experimental evidence shows that humans perceive darker areas of a surface as more distant or deeper [LB00]. This is loosely the basis for the method presented by Langer and Zucker to recover shape on a cloudy day [LZ94]. Khan et al. [KRFB06] successfully demonstrated how, under certain circumstances, limitations in our ability to correctly interpret depth and lighting [OCS05] can be exploited to create plausible synthetic images using a similar *dark-is-deep* approach.

Broadly, our approach to recover meso-scale geometry [GWJ+08] is similar in spirit to the iterative technique of Langer and Zucker [LZ94]. They observe that luminance in the surface depends primarily on a function of the solid angle of the sky that is visible and apply a set of constants and a robust numerical approach to solve depth. We developed a simpler deterministic solution that works in multi-scale image space. Assuming an underlying planar surface, we define a surface meso-structure model that can be approximated as a terrain with hills and valleys with average depth equal to zero (Figure 2.6).

Shape-from-shading approaches are useful to approximate meso-scale geometry from a single image, but they rely on a constraining assumption of uniform reflectance, therefore only mono-chromatic surfaces can be recovered. Estimating the shape of multi-colour surfaces requires separating shading and reflectance, resulting in two *intrinsic images* [BT78]. Intrinsic images have been estimated from a single image using machine learning approaches for classifying gradients in the image into shading

Above-plane surface model applies

Below-plane surface model applies
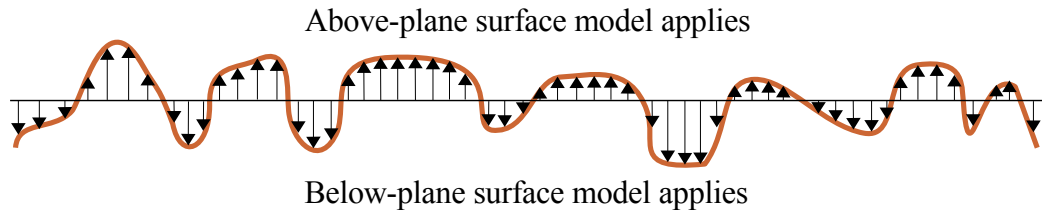
Figure 2.6: *Example of a profile of a textured surface and the separation between the above-plane and below-plane surface models.*

or colour [TFA05]. Systems based on photometric stereo are able to decouple geometry from colour since lighting is known and several images are available. Our system [GWJ⁺08] uses only two images from a fronto-parallel view point, one with flash and one without flash under diffuse lighting conditions (for instance a cloudy day), to solve the shading-reflectance ambiguity. The flash image, after calibration, is a close approximation to the colour image, since the flash also lights the shaded areas, and the shading image is computed as the ratio of the image under diffuse lighting and the colour image. This approach relies on the flash light being located close to the camera lens.

2.2 Recovering Visual Appearance for Outdoor Scenes

When light is incident on a surface, it is reflected back following a spectral distribution and a directional distribution that determine colour and intensity respectively. The spectral distribution is often denoted *texture* and is usually limited to three wavelengths: red, green and blue (RGB). The directional distribution defines the amount of light reflected in every direction for every incoming direction. When this directional distribution is constant for every direction the material is denoted diffuse. Diffuse materials like matte plastic, are also called Lambertian and their reflectance, called albedo, denotes the reflective ratio.

A general model for reflectance is the Bidirectional Reflectance Distribution Function [NRH⁺77], which is normally wavelength dependent and defines both spectral and directional distributions. BRDFs can be represented by a parametric model such as [CT81, LW94, War92] ([Sch94] for a review), which define the complete distribution function with only a few parameters. Extensive research has been carried out to measure accurate reflectance properties from images which is denoted the *inverse reflectometry problem*. In the general case, this requires sampling surface reflectance in

every direction, for every incoming lighting direction at every point, which requires knowledge of the surface geometry, specifically surface normal orientation, and lighting conditions, together with a large number of images. Following this intensive data sampling approach, Dana et al. digitised small patches of complex textures creating a variation of the BRDF called the Bidirectional Texture Function(BTF) [DMP+00]. Marschner et al. [MWLT00] used a curved sample so several directions are sampled at the same time. Efforts to minimise the number of lighting conditions and simultaneously recover 3D shape by extending the photometric stereo principles were presented [Geo03]. Under natural illumination, portable setups using flash lighting can recover reflectance of close views where the flash is still more powerful than the ambient lighting [PCF05, GWJ+08, WG09].

The necessity of controlling lighting conditions limits inverse reflectometry systems to very controlled environments or small regions of the scene. Outdoor scenes require different capture techniques and a different set of assumptions.

2.2.1 Large Scenes Under Natural Lighting Conditions

Large scenes present a level of complexity difficult to approach with traditional inverse reflectometry techniques. Natural lighting is complex and coloured, varies with time and is uncontrollable. Multiple materials may be present and models may have complex geometry.

Two very interesting works managed to recover non-Lambertian reflectance for complex outdoor scenes. Debevec et al. [DTG+04b] presents a full digitisation of the Parthenon using images and a laser-scanned model. The most interesting contribution in this paper is the estimation of the BRDF parameters (Lafortune model [LW94]) using *changing natural illumination*. They designed a measuring device for natural lighting and a portable device to recover some BRDF measurements of representative surfaces. They measure an accessible surface that exhibits a range of colouration properties representative of the site using 83 photographs from different directions under a single illuminant in four different positions. Then they fit the measurements to three BRDFs using a Lafortune cosine lobe model. They infer the most plausible BRDF for a surface point, given its Lambertian colour and the BRDF samples available. Starting with these initial reflectance properties, and a laser scanned model of the building, they use the same iterative process as in [Deb98, BG01] to refine the reflectance properties. The results of this process are impressively realistic. Unfortunately, it requires special equipment, capturing the lighting conditions for every shot of the building and accurate

geometry.

A more applicable method for material recovery of outdoor scenes was presented by Yu and Malik in [YM98]. They use a small number of photographs captured with a hand-held camera to recover photometric properties of buildings in outdoor scenes. They decouple the distant lighting from two main sources, sun and sky. They model the sun as a parallel light and reconstruct manually a low resolution version of the sky and environment at the moment the photographs are taken. They introduce the concept of a *pseudo-BRDF* that separates diffuse and specular components for both sun and sky. For this method to work properly they need at least two images of each face of the building, one with direct illumination from the sun and one without it. A simple 3D model produced with image based interactive modelling software [DTM96], combined with the recovered pseudo-BRDF, produced realistic results. This model lacks self-shadowing effects and it is difficult to appreciate the real benefit of the recovered specular components; it still needs at least two images with different lighting conditions for every façade of a building. Both these techniques recover lighting or environment data in addition to the images.

Recovering reflectance properties for large outdoor scenes requires therefore, a set of different views of the scene registered to a complete geometrical model, and measurements of the lighting conditions for every view. The acquisition and processing of all this data is itself a complex task, so further assumptions are necessary to make this process tractable for non expert users.

A common and reasonable assumption when trying to recover a building reflectance is that most of the interesting construction materials are mainly diffuse. This assumption ignores windows and metallic modern construction materials, but most brick and stone types can be plausibly approximated by a Lambertian material [GWJ+08]. Under this assumption, the inverse reflectance problem is simplified to solve the lighting-reflectance ambiguity, i.e. what proportion of an observed colour is due to reflectance and what due to lighting. Lighting and reflectance can be decoupled with a single observation per point of the material, providing knowledge about the lighting. Unfortunately, capturing the lighting in outdoor scenes requires special equipment [Deb98, STJ+04, DTG+04a] and a second camera synchronised with the one we use to take the material samples.

In cases where illumination cannot be measured because of equipment limitations, some alternatives have been presented. Xu et al. [XGRD06] used the ratio between the green channel of the captured images and the reflected laser intensity to correct all

colour channels. Without the extra information provided by the laser scan, decoupling lighting and illumination is an ill-posed problem. Triccoli and Allen [TA08] presented two algorithms to deal with outdoor reflectance recovery using only image data and without capturing illumination. The first algorithm allows for relighting using the ratio between two images of the same surface under different lighting conditions. For a diffuse surface, such a ratio is texture free, i.e. it cancels out the albedo. This work is similar to the one of Agathos et al. [AF03] and Beauchesne and Roy [BR03] for fusing views of the same object under different lighting conditions. This assumption does not hold in the presence of shadows and fails in the case of pronounced self-shadowing when highly textured surfaces are captured under direct lighting. The second algorithm recovers albedo up to a scale factor. It uses an illumination model for the sky and sun similar to [YM98] based on a point light source for the sun, which is recovered using the time stamp from the camera and a database for sun trajectories in the location of the scene, and a low dimensional spherical harmonic expansion following the work of Ramamoorthi and Hanrahan [RH01]. This method shows the limit of what we can achieve in radiometric terms from the images of the scene only. The problem can only be solved up to a scale factor, needs at least two different illuminations, simplification of the illumination model, acquiring extra information from GPS and solar movement databases, but it does handle artifacts like shadows.

2.3 Appearance Modelling Based on Exemplars

We have seen that modelling realistic and relightable large outdoor scenes such as buildings remains a complex task that requires specially dedicated equipment, is very time consuming and is far from automatic. The aim of this thesis is to explore simpler, more approachable, more automatic methods to achieve high quality results comparable to the ones provided by the state of the art techniques in heritage reconstruction [DTG$^+$04b].

Image-based modelling is a feasible alternative and its relatively low cost allows widespread adoption. However, reconstructing models from uncalibrated images still presents problems to acquire geometry at the level of detail of a laser scan. Automatic multi-view approaches and interactive systems allow the user to recover good low-medium resolution models, but struggle to acquire reliable and realistic texture detail suffering from alignment and resolution limitations. High resolution texture geometry

is important for plausible relighting and view point change, and can be acquired by using controlled lighting, even under natural conditions. These techniques, however, are limited to the power of portable light sources such as flashes, whose effect is limited to 1 to 2 meters. This limits the recovery of detailed geometry to accessible small areas. The same restriction applies to reflectance recovery. Without artificial lighting or special equipment to measure natural lighting, the problem of recovering reflectance, even in its simpler diffuse representation, is ill-posed. Therefore, we are again restricted to partial, attainable samples where we can use artificial lighting.

Here, we explore a novel mixed approach that uses multi-view stereo for recovering gross-scale geometry and an exemplar based transfer system to add meso-scale geometry and reflectance to the model. Using the system described in section 2.1.2 we recover a medium resolution model of the scene. We then recover a unified texture map for this model from the multi-view imagery available using a novel technique. We capture meso-scale geometry and reflectance of relatively small samples of representative materials present in the scene using a simple photometric approach [GWJ$^+$08] that only requires two images of a fronto parallel view to recover depth and albedo. Identifying regions of these materials in the global texture map and associating them to the exemplars, we are able to transfer albedo and geometric detail to the gross-scale model. The complete process produces a model comparable, in level of detail, to those recovered with laser scanners, with modest user interaction, easy data capture process, and without the necessity of expensive equipment.

To our knowledge, the idea of transferring geometry and reflectance properties from exemplars is novel and presents a new approach to solve the intrinsic radiometric problems of modelling large outdoor scenes from photographs. Our transfer techniques are totally image-based and take inspiration from the texture analysis, synthesis, and transfer, particularly from the exemplar-based methods, which have proven to be effective for both synthesis and transfer [WLKT09, HJO$^+$01].

2.3.1 Texture Synthesis and Transfer

Texture can be defined as an image containing repeating patterns. Natural textures may contain interesting variations or imperfections which are referred to as randomness. Depending on the amount of randomness, textures can be classified within a continuum from regular (tiled floor) to stochastic (sand beach). Texture synthesis solves the problem of creating an arbitrary sized image of a certain texture from a smaller sample. On the other hand, *texture transfer*, first defined by Efros and Freeman [EF01],

takes two images – the source image and the target image – as input and modifies the target image to acquire certain characteristics of the source image. The objective of texture synthesis is to generate textures that are similar to the source texture. *"A clear criterion of success exists in texture synthesis: the result has to look like the input"* [Ash03]. However, establishing formally what "look like" means is difficult. Most techniques are considered to perform correctly when users evaluate the results as plausible. Some efforts to quantitatively evaluate texture synthesis algorithms have been carried out recently [ZFG08]. Texture synthesis and transfer share many of the same challenges and several texture synthesis algorithms have been also used for texture transfer [EF01, Ash01, Ash03]. Texture transfer is normally used for artistic purposes and formal evaluation has not been studied.

We take inspiration from two different techniques that were successfully applied to texture synthesis and transfer. The first one is histogram matching [HB95]. Heeger and Bergen demonstrated that by matching the statistics (histograms of the response of two images to filters at different scales and orientations) of two images, they could synthesise very plausible results for stochastic textures (see [ZWM98] for a theoretical justification). However their approach failed for synthesising more structured textures. We apply histogram matching as a transfer technique and use it to transfer albedo and meso-structure. The source image is the albedo or shading map to be transferred and the target image is the global texture. We show that we can transfer reflectance and meso-scale geometry, generating plausible results.

The second transfer method is also inspired by the texture synthesis literature. In order to correctly synthesise more regular textures, methods started using neighbouring information to gradually generate texture by finding the best pixel by comparing its neighbourhood with the neighbourhood of the already synthesised pixels [EL99, WL00a]. These algorithms are simple to implement, but are slow since searching for the best candidate (*Nearest Neighbour Search* (NNS)) is a costly operation. Ashikhmin explored the concept of coherence (neighbouring pixels in the input are likely to be found close together in the output) to improve the performance of the algorithm [Ash01]. This work also allows for a user defined colour map to control the synthesis. Efros and Freeman use a real image to define a correspondence map instead, defining texture transfer [EF01]. Hertzmann et al. explored this concept further, and proposed a technique to automatically create correspondence maps, which they called *image analogies* [HJO+01]. Image analogies take three images as input: a source image, a filtered version of the source image, and a target image. It uses a similar algorithm

to [WL00a, Ash01] to find correspondences between the source and target image, generating a correspondence map, but storing the relative offset between both images. Using this correspondence map together with patches from the filtered image, it is able to create an analogous filtered image of the target image. Building upon this prior work, researchers have proposed methods to synthesise geometric detail [BIT04], geometric texture [ZHW+06] and height fields for terrain synthesis [ZSTR07].

We use the concept of Image Analogies to transfer albedo and meso-scale geometry. We implemented several optimisations and extensions of the original technique, accelerating the NNS [BSFG09] to allow for high-resolution texture maps, and to handle the specific case of transferring material characteristics.

2.4 A Novel Mixed Image-based Approach for Outdoor Scene Reconstruction

Having presented the context and definitions necessary for the rest of this thesis, we present the global structure of our modelling system. Figure 2.7 shows the work flow and different stages of the process. Orange boxes are techniques developed and implemented for this thesis. Grey boxes are stages where we use available software, and white boxes are the processes where manual intervention is needed. Note that only cleaning the gross-scale mesh and some segmentation options are not fully automatic. The complete colour code is the following:

- **Grey**: Other available software used.

- **Orange**: Developed and implemented for this thesis.

- **White**: Manual process.

- **Turquoise**: Intermediate data.

- **Pale Yellow**: Input data.

- **Bright Yellow**: Output data.

We start from two kinds of image input data. One or several sequences of images from different points of view of the whole building are used to acquire the gross-scale model, reviewed in section 2.1.2. The gross-scale model is flattened into a plane and used, together with all the images, to create an optimal texture map. We describe this
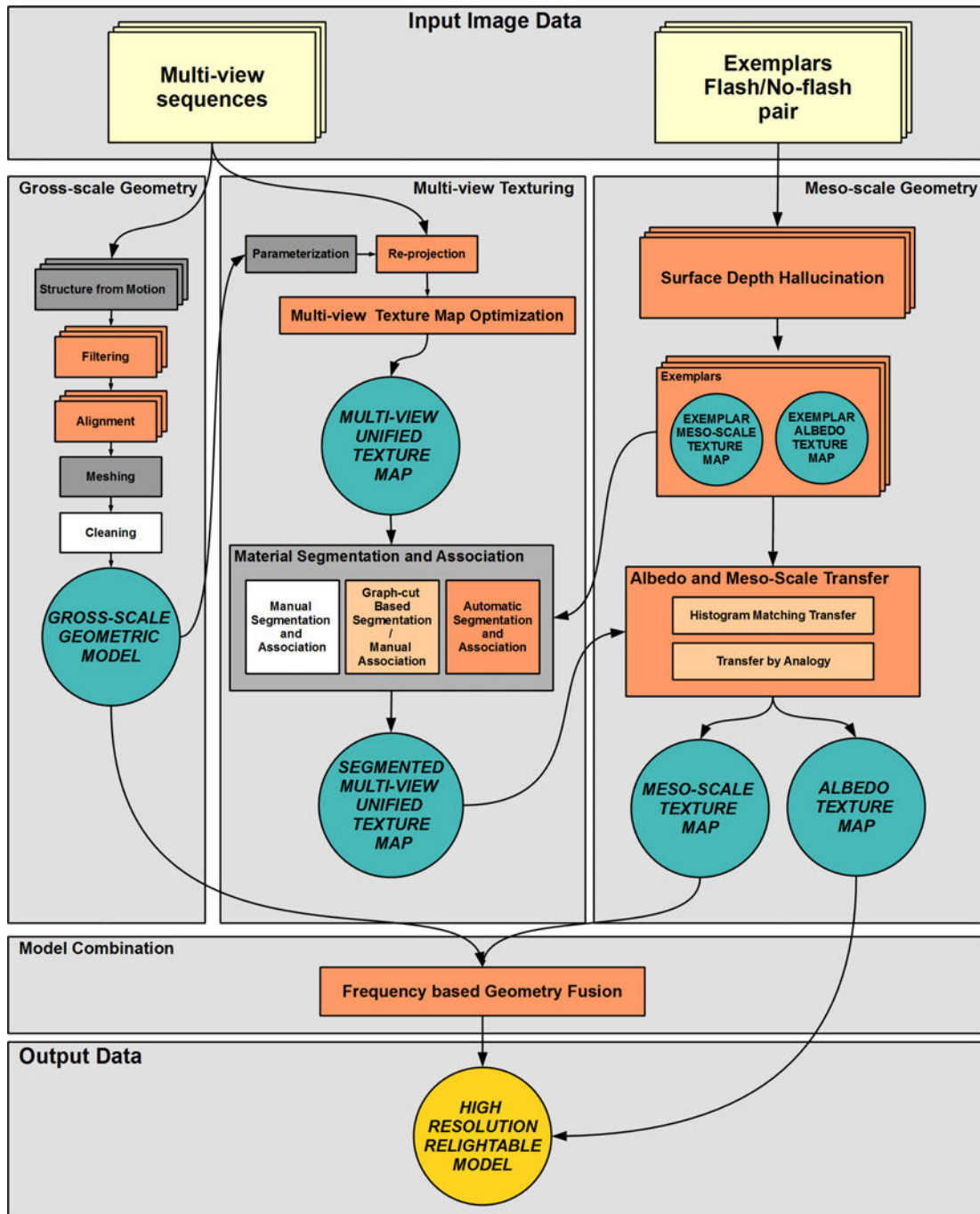
Figure 2.7: *Schematic of our Mixed Image-based Modelling System.*

process in Chapter 5. Exemplars of different materials present in the scene are captured and processed using surface depth hallucination [GWJ+08]. These exemplars will be used to transfer meso-scale structure and albedo to the surface of the model. Chapter 3

describes the capture process and presents and evaluates the transfer techniques. Segmentation of the global textures and transfer of characteristics for multiple materials are covered in Chapter 4. Finally, the process of combining together gross-scale and meso-scale models is presented in Chapter 5.

Capturing and Transferring Texture Appearance and Meso-structure

THIS *chapter investigates the capture and transfer of appearance and geometry from images. It reviews a method to capture both depth and albedo from small portions of globally flat surfaces from a single point of view. It also investigates the relationships between appearance under diffuse lighting conditions, reflectance and depth, and presents and evaluates two methods to transfer these material characteristics to different samples of the same materials.*

Texture has several definitions. Here the term texture denotes the characteristics of a surface, and is represented by an image or a set of images that are mapped to a 3D geometrical model. The image or images that represent the texture are denoted *texture maps*. Texture is a very powerful tool to add important characteristics to 3D models, for example to provide geometrical detail, colour information, or realistic appearance, and therefore is a crucial aspect of 3D modelling.

There are three different approaches to the creation of textures: capture, artistic generation, and synthesis. Capturing a texture can be as simple as taking a picture. Unfortunately, a simple picture is usually not a good enough representation of the texture since it is illumination and view point dependent and therefore it is only valid under the lighting conditions and view point under which the picture was taken. Since capturing an appropriate texture map is not always possible, artistic generation and synthesis are also required. Texture synthesis is probably the most active area of research related to textures and some problems with this remain unsolved.

Different characteristics and different types of texture present different problems when synthesising and capturing images. We focus on two essential characteristics of the texture: *albedo* and *meso-structure*. An albedo map represents the diffuse reflectance of the surface. It is, therefore, a colour map that defines the ratio of reflected light to the incident light, under the assumption that this is equal in all directions.

Unlike a photograph of the surface, the albedo can reproduce correctly the colour of the surface under novel lighting and viewpoints since it does not contain shadows or colour contributions from the light sources. The variations in appearance due to self-shadowing effects require meso-structure to be realistically reproduced. Meso-structure refers to the relief or texture geometry and is represented by a depth map.

These two texture characteristics do not represent a physically accurate representation of all materials, but Glencross et al. showed, through evaluation with human subjects, that this model provides enough information to produce perceptually plausible relightable models [GWJ+08]. Figure 3.1 shows renderings under novel lighting using this albedo-plus-meso-structure model, using only the albedo map, and simply mapping a photograph of the textured surface under natural lighting. Figure 3.1(b) shows how without meso-structure the result looks flat as if it were painted on a plane. Using the original photograph (Figure 3.1(a)) provides a richer appearance but shadowing and colour are not consistent with the lighting. The result using the complete model in figure 3.1(c) presents the appropriate shading and colour appearance.

(a) Original Image of the material under diffuse lighting conditions.   (b) Recovered albedo map.   (c) Recovered meso-structure.

Figure 3.1: *An example of a material recovered using Surface Depth Hallucination. Albedo and meso-structure provide a realistic appearance model for relightable models.*

As presented in chapter 2, capturing albedo and meso-structure for a large structure, like the façade of a building, is not always possible or practical because of accessibility issues, necessity to control or measure lighting, and it can be a very time consuming process for large scenes. Texture synthesis is an alternative approach that can complement texture acquisition either automatically or interactively. The repeating nature of texture patterns allows us to create large textures from an input exemplar. For example a 'brick' pattern can be used to cover a large wall by repeating the pattern across the surface. Real textures, contain irregularities and interesting variations, referred to as randomness, which are difficult to synthesise.

Depending on the amount of randomness, textures can be classified within a continuum from regular (tiled floor) to stochastic (sand beach). Textures in the middle of this spectrum are the most difficult to synthesise realistically. Texture synthesis, particularly for near-regular textures, has received a lot of attention in the last decades in the research community, and many advances have been achieved [WLKT09]. Synthesising albedo and meso-structure using a captured exemplar is therefore a possibility and approaches exist to produce colour texture [Ash01, EF01, Ash03] and geometric texture [BIT04, ZHW+06, ZSTR07].

Although texture synthesis may produce a similar texture, the result does not necessarily correspond to the real texture of an existing surface. Since capturing the complete information is not always possible and synthesis has some limitations, we aim to combine both capture and synthesis and propose to use a *transfer* approach. Transfer can be understood as an exemplar-based guided synthesis. We capture albedo and meso-structure of a representative portion of a material which constitutes our exemplar. Then, a new image for the complete area containing this material, captured under certain lighting conditions, is used to guide the synthesis. Our goal is to produce the albedo and meso-structure maps corresponding to the guide image, which contain the original variations of the real surface. Texture transfer techniques have been used for artistic purposes [Ash01, EF01, HJO+01] but, to our knowledge, they have not been used to transfer physical properties of texture. This chapter presents the problem of capturing and transferring albedo and meso-structure, analyses the specific relationships between these texture characteristics and appearance, and proposes and evaluates two image-based techniques to perform the task: *Histogram Matching* and *Transfer by Analogy*.

We first review our exemplar capture technique, *Surface Depth Hallucination* (**SDH**). This technique was developed in our department (with the author of this thesis as contributor), but is not claimed as a contribution of this thesis. However it is crucial to understanding the motivation for the chapter and rest of the thesis. The conclusions presented in the evaluation (included in Appendix A) examine it as a robust technique to capture perceptually plausible models of different textures.

The final aim of our transfer techniques is to provide perceptually realistic appearance to more general scenes involving complete buildings and other large structures. In this chapter, we focus on evaluating the capabilities of our algorithms to transfer the proposed texture characteristics. We analyse their transfer capabilities and theoretical limitations, and evaluate their behaviour with real data. To do so, we test the

techniques on small samples of the materials where we can capture ground truth data against which to compare the results of the transfer techniques. We use exemplars captured with the perceptually validated SDH technique as ground truth data. Finally we discuss their advantages and disadvantages, and their application in the context of 3D architectural reconstruction which is further discussed in the following chapters.

3.1 Capturing exemplars: Surface Depth and Albedo Hallucination

(a) Original Image of the material under diffuse lighting conditions.

(b) Recovered albedo map.

(c) Recovered meso-structure.
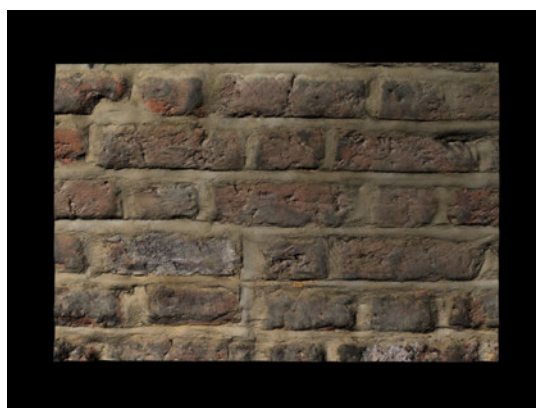
(d) Complete Recovered Model rendered under novel lighting conditions.

Figure 3.2: *An example of a material recovered using Surface Depth Hallucination.*

Recovering surface characteristics of textured surfaces from images is a challenging problem. Recently, we presented a method that can recover sufficient characteristics to be able to reproduce realistically surfaces covering areas of approximately one square meter, which we named *Surface Depth Hallucination* [GWJ⁺08]. The main advantage of this method with respect to previous approaches is that it only needs one point of view and only two images during data capture, one with flash and one without. These images are taken under natural lighting conditions but with the single restriction that no direct sunlight hits the surface, so that images must be captured, for example, on cloudy day or in a region in shadow.

This method was evaluated by users who rated renderings of the reconstructed models as real images, proving the perceptual plausibility of the results. Participants were asked to rank each image from 1 to 5, corresponding to their certainty that the image they were viewing was an untouched photograph. On this scale, we define 1 as definitely synthetically generated, 5 as definitely an untouched photograph, and 3 as undecided. On our rating scale a value of above 3 suggests the image is more likely to be a photograph than synthetic. Although people rated photographs higher, re-lit images were rarely dismissed as artificial, and equivalent photographs were not always recognised as real. Importantly, around 15 out of 20 participants gave our synthetic images average ratings above 3, leading us to conclude that our renderings compare very well with photographs. This was further supported by participants commenting in post study debriefing, on the difficulty in determining which images were synthetic. More details about the perceptual study are presented in the Appendix A.

The process is overviewed in Figure 3.3 and consists of two stages. First is the disambiguation between albedo, shading, and lighting which delivers an albedo map and a shading map. Second is the calculation of depth from the shading map which delivers a depth map that can be used to generate a geometric model and a normal map. We use this technique to capture exemplars of textured surfaces.

We assume our surface can be plausibly represented as a height field, whose underlying material is approximately Lambertian and opaque, with average reflectance, $2\% < \rho < 70\%$. The input to our process is a diffuse-lit/flash-lit image pair [ED04]. Subtracting the diffuse-lit image from the flash-lit image gives a reasonable estimate of albedo, and the ratio between our diffuse-lit image and albedo provides a usable estimate of diffuse shading for depth estimation. The final models can be rendered under novel lighting conditions and viewpoints.
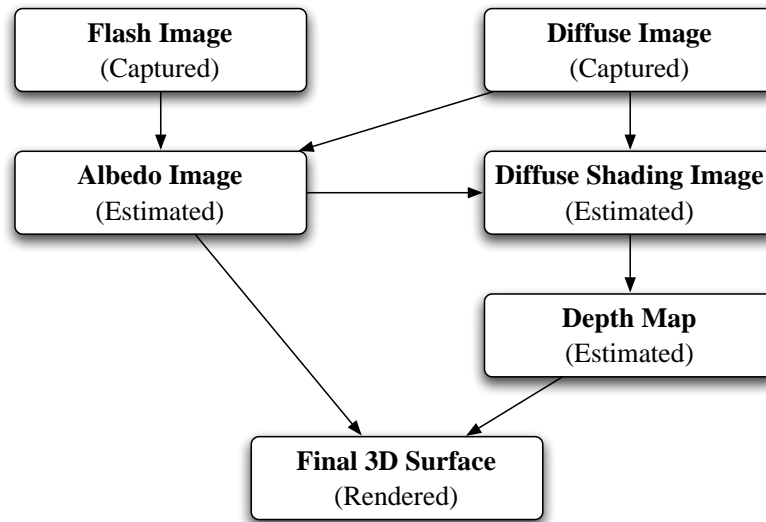
Figure 3.3: *Flow chart showing the steps in of the exemplar capture process.*

3.1.1 Image Capture

To capture our input images, we employ a standard digital SLR camera mounted on a tripod, and an attached strobe to achieve good alignment. Our method requires that we capture a sample of the textured surface without global curvature, as might be found on a wall or floor.

First we capture an image under indirect illumination (i.e., overcast skies or shadow). We call this the diffuse-lit condition. A second photo is taken from the same point with the flash fired at full power. Ideally, the flash should be mounted as close to the camera lens as possible in order to minimise shadows, though the images shown in this thesis

(a) Photograph of a brick wall taken in shad-owed daylight conditions.

(b) Flash-lit photograph of the brick wall.

Figure 3.4: *An example input photograph pair.*

were all taken with a standard flash mount. See Figure 3.4 for an example input image pair.

3.1.2 Albedo Map and Shading Image

The first stage in our method requires estimation of albedo and diffuse shading. This results in an albedo map and a shading map that will be used subsequently in the transfer process. Pixel values in the diffuse-lit image I_d are subtracted from our flash-lit capture I_f, and we divide the result by pixel values in the flash calibration image I_c taken of a white Lambertian surface at a similar distance and aperture. This yields approximate reflectance values at each pixel, simultaneously correcting for vignetting, fall-off, and the global cosine factor.

We apply a daylight white balance that provides a good match to the flash, therefore image subtraction results in a good colour balance in our albedo image, as shown in Figure 3.5(a). In cases where flash shadows are present, we also apply a simple thresholding and neighbour-filling technique that copies detail from the flash-lit areas [PSA$^+$04].

To compute the diffuse shading image, we take the ratio of the diffuse-lit condition over the albedo at each pixel. This can result in a colour cast due to skylight or cloudy illumination, but our depth estimation method uses only the luminance channel. A computed greyscale shading image for our brick path is shown in Figure 3.5. The depth estimation method described in the following section assigns a height of 0 to a pixel intensity of 0.5, so we normalise our shading image to this mean value.

(a) Derived albedo map.   (b) Diffuse shading image.
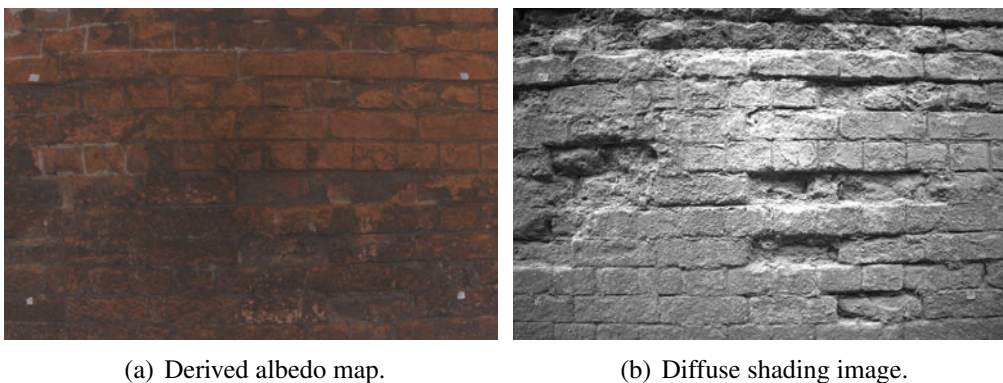
Figure 3.5: *Example albedo map and shading image generated from the photographs in Figure 3.4 of the brick wall.*

3.1.3 Depth Estimation

Estimation of depth from shading is a popular problem in computer vision [ZC91].
The Langer and Zucker [LZ94] method is designed to recover shape from shading
on a cloudy day by using an iterative ray-tracing approach. Instead we developed an
approximate solution that works entirely in image space and yields a direct estimate of
depth at each pixel.

   We define a surface meso-structure model that can be approximated as a terrain
with hills and valleys with average depth equal to zero. We therefore begin by devel-
oping two local models to approximate these different types of relationships between
meso-structure depth and shading. We derive our below-plane shadowing model by
approximating pits in the surface as cylinders (Figure 3.6(a)), and for the above-plane
model, we approximate surface protrusions as hemispheres (Figure 3.6(b)). These
models are derived such that an above-plane linear model is matched to a below-plane
quadratic model at a tangent point, creating the smooth piecewise function plotted in
Figure 3.6(c).

(a) Cylinder pit model. (b) Hemisphere protrusion model. (c) The relationship between aperture and shading factor in our model. The dashed line shows the unused extensions of each model.
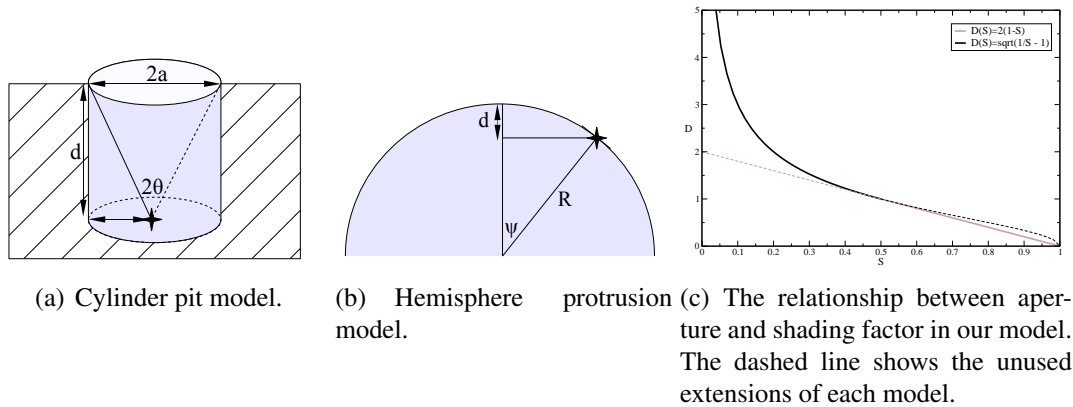
Figure 3.6: *Model to approximate shading of pits and surface protrusions.*

   As noted earlier, our depth estimates are conservative. First, we ignored albedo
to simplify our analysis. Second, we approximated indentations in the surface as pits,
where a crevice model might be more appropriate in some cases. We therefore ap-
ply a user-selected, uniform scaling factor to each depth map to compensate for this
and achieve an acceptable visual match to the original surface appearance. The depth
computation is perform in a multiscale process by estimating the depth in the different
scales and adding them up, producing a complex smooth signal.

3.2 A Transfer Approach

Surface Depth Hallucination allows us to capture approximations, in the form of albedo
and depth maps, of texture reflectance and meso-structure. This provides enough infor-
mation to simulate perceptually valid models of many types of building textures under
different lighting conditions and view points. Unfortunately, this technique is limited
by the necessity of lighting the surface with a flash, whose standard working distance
is one to two metres, so only patches of approximately 1 square meter can be captured
in a shot. Furthermore, accessibility to high areas of the façade or rooves is highly im-
practical. Capturing texture for the whole building without the limitation of artificial
lighting is a simpler problem. We propose to capture a complete texture map under
natural lighting conditions and use this map as a guide to transfer albedo and meso-
structure from the representative exemplars captured with SDH. Before we detail the
transfer process for full façades, we present and evaluate two image-based transfer
techniques inspired by the texture synthesis and transfer literature [HB95, HJO+01]
using small patches of textured materials. The rest of this section defines formally the
material transfer problem and the relationships between albedo, meso-structure, and
appearance under natural illumination.

3.2.1 Material Transfer: Definition

We state the problem of material transfer in figure 3.7. A material M is defined by an
exemplar A constituted from three maps: an ambient map (A_{am}), a shading map (A_s),
and an albedo map (A_a). Now consider another sample B of the material M, for which
only the ambient map is being captured. The aim is to synthesise a shading map (B_s)
and an albedo map (B_a), from B_{am} and the exemplar A. These maps will be used to
produce a full model using the depth estimation presented in section 3.1.3.

Transferring Albedo and Shading from Ambient Lighting Capture

According to this definition, we assume that albedo and shading can be approxi-
mated/synthesised from a capture under ambient diffuse lighting conditions. Essen-
tially, that involves resolving the ambiguity between albedo and shading, which is an
ill posed problem from a single observation. The colour appearance can result from
multiple combinations of surface reflectance, lighting, and meso-structure. Assuming
that the same material has similar characteristics, we use an exemplar where this am-
biguity has been solved to help in making a good guess about both albedo and shading
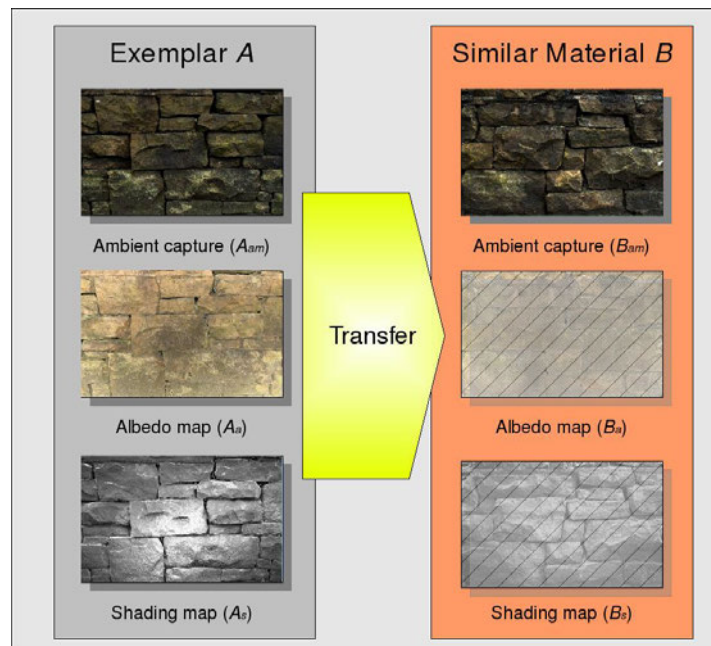
Figure 3.7: *Transfer schematic.*

of the new image. We follow with an analysis of the relationships of albedo, shading and ambient capture.

The albedo and the ambient images have two main differences. The ambient capture contains shading effects, thus some pixels will be in shadow and will appear darker than their true albedo. The second difference is the lighting conditions: the albedo is the reflected light from the full spectrum (white illuminant), but the ambient map has some colour shift due to ambient illumination which depends on the environment around the capture, the sky, and the sun. The albedo map is always brighter than the ambient capture. Inferring the albedo map from the ambient map will require therefore a shift of the value of the pixel to a brighter level.

The shading map and ambient map have a different relationship. In order to compare both maps we transform the ambient map to greyscale, using the luminance Y which is a linear combination of the Red, Green and Blue channels in the form $Y = 0.2125R + 0.7154G + 0.0721B$ following ITU-R Recommendation BT.709. The shading image results from computing the ratio between the ambient map and the albedo map, which is then normalised to have an average 0.5, meaning a completely flat surface would have a constant value 0.5 across the whole image. The grey scale ambient map depends on the albedo, so dark and light areas may be due to shading, albedo, or both. In general, the per pixel relationship between the shading map and

54

the ambient map is arbitrary, with the exception that a completely white pixel in the
ambient map ($Y = 1$), cannot have a shade value smaller than 0.5, since it cannot be
in shadow.

In the following sections we introduce two transfer techniques: *Histogram Matching* and *Transfer by Analogy*. We analyse the theoretical transfer capabilities of both
techniques. Then, we show results of the transfer with real data and evaluate its performance. Finally, we suggest future work for material transfer.

3.3   Histogram Matching

(a) Histogram Ambient map(Y)   (b) Histogram Shading map   (c) Matched Histogram

(d) CDF Ambient Map(Y)   (e) CDF Shading Map   (f) Matched CDF
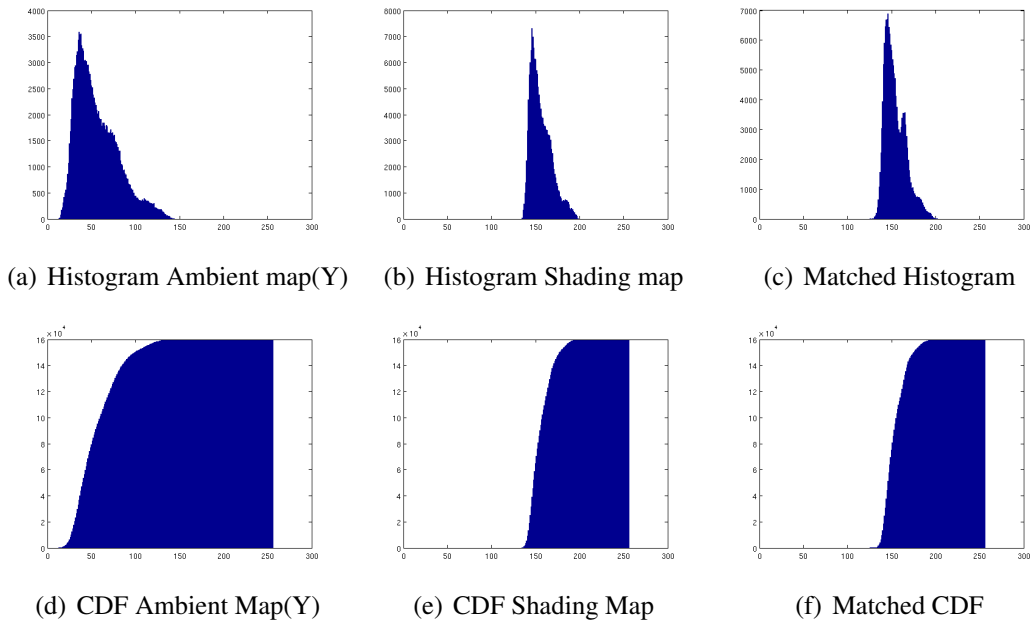
Figure 3.8: *Histogram Matching works by matching the Cumulative Distribution Function(CDF)*

The histogram represents frequencies of the values in an image. It is normally represented graphically by adjacent rectangles representing the count of instances falling
in every bin (Figure 3.8(a)). Histograms are simple descriptors of images, but possess interesting properties that make them very effective to characterise images, for
example for data retrieval [YJL$^+$08]. *Histogram Matching* [HB95] is a technique that
modifies the shape of the histogram of an image to match the shape of a second given
histogram. The basic assumption behind Histogram Matching as a transfer technique
is that two samples of the same material are statistically similar. By matching the

statistics of the albedo and the shading maps we create a new sample of the material that will have similar statistics and consequently similar appearance. This idea was explored for synthesising stochastic textures [HB95] and detail in textures [IBG03] by matching the histogram of noise patterns to a texture sample. Here, we extend this idea to pseudo-regular textures, and to transfer albedo and shading patterns.

The histogram matching algorithm is illustrated in figure 3.8. Having two histograms 3.8(a) and 3.8(b), we want to match the source histogram 3.8(a) to the shape of the target histogram 3.8(b). The algorithm first computes the *Cumulative Distribution Function* (CDF) (Figures 3.8(d) and 3.8(e)) of the target histogram, then it takes the counts in the bins in ascending order of the source histogram to reconstruct the same cumulative distribution.
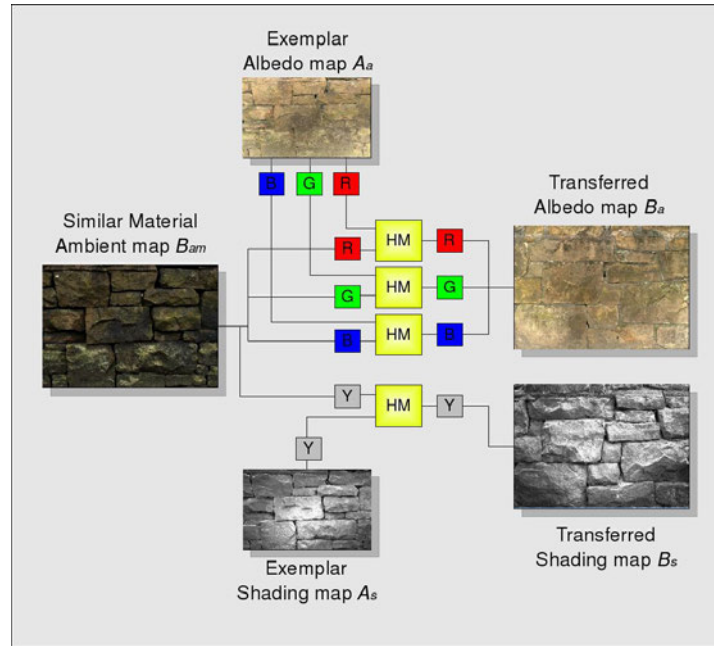
Figure 3.9: *Transfer schematic.*

The material transfer process using Histogram Matching is illustrated in Figure 3.9. Given a complete exemplar A, with maps A_{am} (A_{am} is not shown in Figure 3.9 because it does not contribute to this transfer process), A_a and A_s, and a different ambient map of a similar material B_{am}, we use histogram matching to create B_a and B_s. Albedo is transferred by matching the histograms of the R, G, and B channels separately and shading is transfered by matching the histogram of the luminance channel Y.

Glencross et al. [GWJ$^+$08], presented some preliminary results showing that using

this technique we can generate approximations to albedo and surface detail for similar surfaces obtaining perceptually plausible results that were validated in a user case study. These results can be seen in Appendix A. In the following section we analyse theoretically and empirically the performance of this technique, to understand in detail what is capable of transfer and what is not.

3.3.1 Analysis of Transfer Capabilities: Histogram Matching

In order to analyse Histogram Matching as a transfer technique, we ran the following experiment. Given a complete exemplar A, with maps A_{am}, A_a and A_s, we took $B_{am} = A_{am}$. Therefore A and B are samples of the same material and statistically similar (actually identical), as required by transfer technique. Histogram Matching is a stochastic process so the result of the transfer is not trivial and A_{am} and B_{am} being the same image does not guarantee a good transfer. A perfect transfer will produce maps B_a and B_s identical to the captured A_a and A_s respectively. Since the Histogram Matching process is performed per channel, we analysed the behaviour of a single channel (the luminance channel) which coincides with analysing the transfer mechanism of the shading image. The conclusions can be extrapolated to every channel of the albedo image.

| (a) Ambient map | (b) Shading map | (c) Monotonic H^2 |
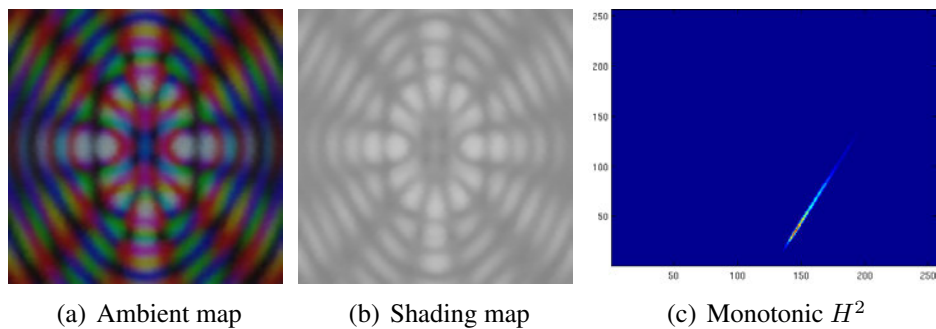
Figure 3.10: *Two dimensional histogram of a synthetic exemplar*

The algorithm matches the histograms in ascending order. This is a key aspect of this process since it has an important consequence: if a pixel p is darker than a pixel q in the ambient image, p will always be to be darker or equal than q in the transfered shading image. The relative value order between pixels must be the same in the ambient and shading images, since histogram matching is not able to change this order. The histogram matching process sets the brightness and the contrast of the

image to match the ones of the exemplar, but does not change the relationship between
the ambient appearance and the shading.

Thinking about it graphically, we plot a two dimensional histogram H^2 (Figure 3.10),
where the possible values of the channel Y of the ambient image are on the y axis, the
possible values of the shading image are on the x axis. A material will be perfectly
transferable by histogram matching when this histogram define a monotonic curve,
meaning that the value relationship always increases. This condition is satisfied for
example for a material with a constant albedo map where all the variations in the ap-
pearance are due to shading. In this case, the H^2 is perfectly monotonic. In the case
of transferring the albedo map, if the shading is constant (a flat surface) the histogram
matching only needs to compensate for the difference in the illuminant. Under diffuse
lighting, the illumination is constant for the whole image, and will affect the albedo as
a colour shift that can be seen as a constant multiplication per channel, which is a linear
relationship that is appropriately corrected with this method. Figure 3.10 shows a syn-
thetic more complex case, where the albedo is not constant, but shading and ambient
obey a monotonic lineal relationship.

Real textures, however, rarely have a H^2 that defines a perfect monotonic curve
as can be observed in Figure 3.11. In general shaded pixels will appear darker in the
ambient image than some unshaded darker albedo pixels. Similarly some dark albedo
pixels will look dark in the ambient image even if the shadowing value is bright. In
general, some shading will remain in the albedo map and some colour information will
be misinterpreted as shadowing on the shading image. The accuracy of the transferred
maps depends on the distribution defined by its H^2.

Besides this limitation our user study presented in Appendix A showed that the
method produces perceptually valid results when rendering the histogram matched
model created form transferred shading and albedo maps. This is due to the nature
of the resulting artifacts, the shape from shading algorithm, and also due to limitations
in the human perception system to evaluate depth from a single point of view [Ram88].

The artifacts in the albedo are caused mainly by remnants of shading effects, mean-
ing darker pixels in shaded areas. These areas are often shaded in the final renderings
producing over shaded areas that are consistent with the current shading being difficult
to detect by the observer. On the other hand, the shape from shading model combines
a linear above plane model with an exponential below plane model. Thus, differences,
and therefore inaccuracies, in the shading image in the below plane region (values un-
der 0.5) have a bigger impact in the final depth map than similar differences in the
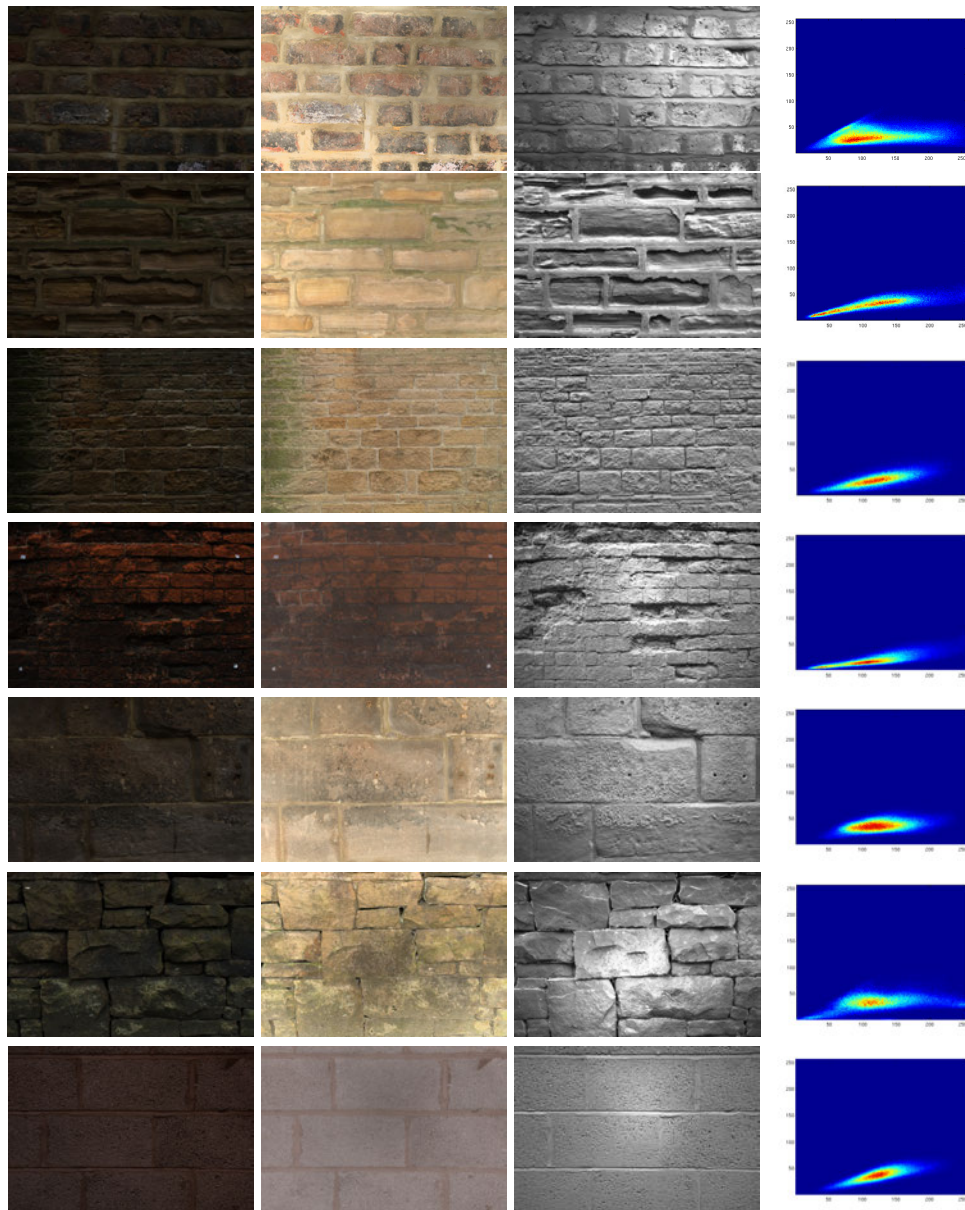
Figure 3.11: *Five of the exemplars used for evaluation. Ambient map; Albedo map; Shading map; H^2. The colour in H^2 means the frequency of each value: red high frequency, blue low frequency.*

above plane. Since the errors in the below plane model produce shaded areas and are therefore rendered as dark pixels, these errors are more difficult to appreciate.

Finally, from a single view point, as in our examples, and without previous knowledge about the surface, an observer cannot resolve the shading-albedo ambiguity further than sensible explanation [Ram88]. Importantly, this means the resulting depth

map and albedo map are required to be a plausible guess that can be accepted by a
user rather than an exact result. The perceptual validation of histogram matching ren-
derings showed that the algorithm produces plausible explanations for both albedo and
meso-structure.

3.3.2 Experimental Evaluation

We used a total of 21 exemplars and a synthetic test exemplar for our evaluation. We
then created an albedo map and a shading map using the histogram matching tech-
nique. We ran two experiments. First we used the same ambient map for A_{am} and
B_{am}. This way we make sure that the both materials are exactly the same. Therefore,
this experiment evaluates the transfer capabilities of the histogram matching algorithm.
The theoretical analysis predicts that the accuracy of the result will depend on the H^2
defined in section 3.3.1. The second experiment uses pairs of exemplars of similar
materials so $A_{am} \neq B_{am}$. This experiment aims to evaluate the performance of the
method when samples of the material are not identical but similar.
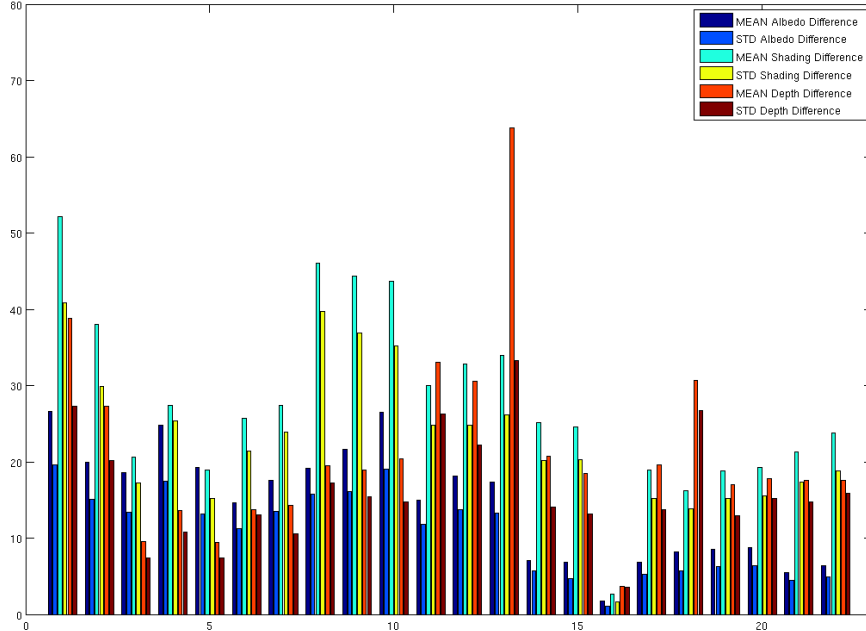
Figure 3.12: *Transfer schematic.*

We evaluated numerically the transfer results, computing the difference between
the albedo, shading and depth maps captured with maps computed using Histogram
Matching. Figure 3.12 shows the global difference between maps in the form of the

mean value and the standard deviation of the error distribution. In the case of the
albedo we used the average of the difference of the three channels. Values in the image
space are in the range 0..255.

In general we see that the error of the estimated albedo is lower than the error in
shading. 38 percent of the exemplars evaluated presented a mean error below 10 points,
which means a 4% difference. 42 percent presented a mean error between 10 and 20
points, i.e. between 4% and 8%. Ninety percent of the exemplars present a mean error
in the albedo below 10% (25.6 points over 255).

Differences in the shading maps are higher. Most exemplars have an error of around
10% with some cases rising to a mean difference in value of 20% (51 points). However,
in most cases the error in the shading maps is reduced when computing the depth map,
especially in those where the shading map presented higher differences. We observed
a case where this effect was strong (exemplar 8) and a case where the opposite also oc-
curs (exemplar 13). Figure 3.13 shows false colour images of the differences between
albedo, shading, and depth maps. Note that these images are scaled for visualisation
purposes and follow the value scale on the right side of each image.

(a) Albedo Difference (b) Shading Difference (c) Depth Difference

(d) Albedo Difference (e) Shading Difference (f) Depth Difference
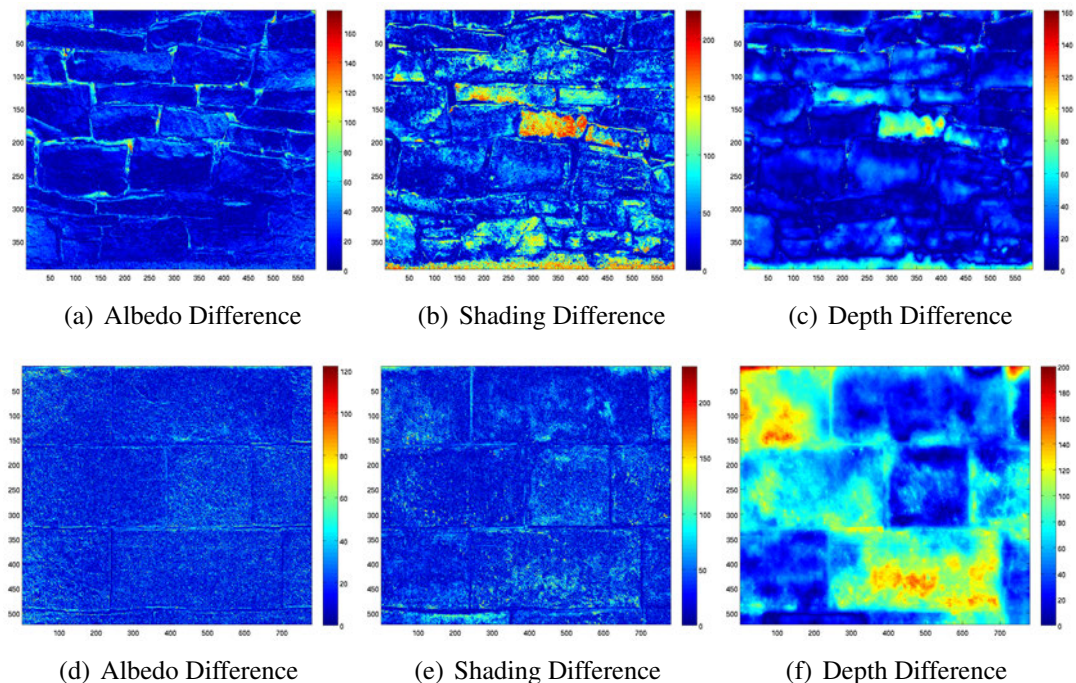
Figure 3.13: *Error in Albedo, Shading, and Depth, for two exemplars.*

Exemplar 8 in Figure 3.13(a,b,c) shows localised high differences in shading with

difference values around 200 (80%). This extreme case happened because this particular area is a protrusion that is interpreted as an indentation because of its black albedo. Fortunately, when calculating the depth map, these differences reduce to 40%. This is due to a smoothing effect produced by the multi-scale nature of the depth estimation that encourages smooth transitions in depth.

The opposite case happens with exemplar 13 in Figure 3.13(d,e,f). In this case the shading images are similar globally and the differences are smaller and more regularly distributed. The same smoothing effect, in this case, causes larger areas with depth disparities. Since the depth map in this case is close to flat, the final depth differences are large, with areas exceeding 50%.

The depth estimation process described in section 3.1.3 contains a scale factor which we fixed for all our experiments, that can be controlled by the user to adjust the roughness of the texture. Although we focused our research on automatic methods, an interactive interface where this factor could be modified by the user with real-time feedback would improve the depth estimation when the automatic result is not satisfactory. We also suggest as future work, to study the customisation of this factor in different scales.

To finalise our analysis, we take exemplar 1 ($Brick1$) where the errors produced by this technique are higher than the average and study the causes compared to those predicted in the theoretical analysis.

3.3.3 Specific Analysis of the Exemplar $Brick1$.
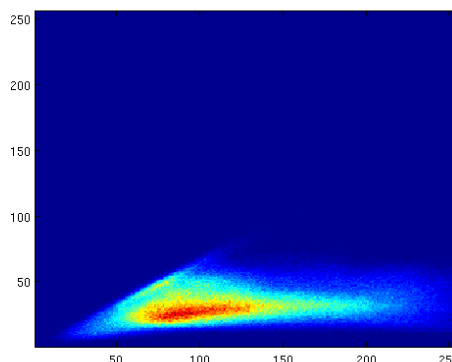
Figure 3.14: *Complete study: H^2 Histogram Exemplar $Brick1$*

The H^2 histogram for the shading map of the $brick1$ exemplar is shown in Figure 3.14. The vertical axis y represents the values of the greyscale image resulting

from the desaturation of the ambient map. The horizontal axis x represents the values of the shading map. Both maps are scaled to the range 0 to 255. The image shows the number of pixels having the corresponding values in both axes. The false colour mapping assigns red for high frequencies and dark blue for low frequencies. First thing to notice is that the distribution is somewhat sparse and does not follow a monotonic curve which we established as a condition for a perfect transfer. This means that the Histogram Matching process will behave poorly for the shading transfer of this exemplar and will incur artifacts, which makes this case interesting for analysis.

Exemplar Brick1:Shading and Meso-structure Artifacts

Differences between the shading images are high due to the high dispersion of H^2 in Figure 3.14. The dark areas due to the colour of the brick are misinterpreted as shaded areas and therefore as indentations; similarly light coloured areas in the albedo will be interpreted as protrusions. Figure 3.15 shows renderings of the captured and the histogram matched meso-structure models. We also render the full model, including albedo, under novel lighting and view point. In addition to the depth comparison, we can compare the changes in orientation of the surface, which is encoded in the normal map computed from the depth map.

(a) Captured Meso-structure and Normal Map

(b) Histogram Matched Meso-structure and Normal Maps

Figure 3.15: *Captured and Transferred Meso-structure. Large artifacts due to uncorrelated appearance and shading appear, however high frequency features are correctly estimated.*

We extract interesting conclusions from the visual evaluation of the renderings in Figure 3.15. The depth estimation is wrong, but still coherent with the texture and
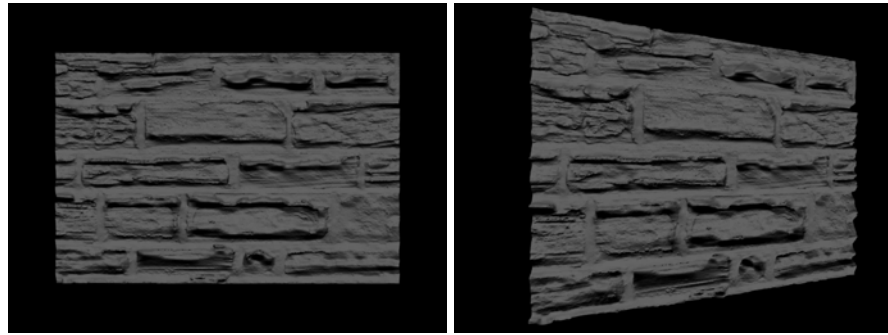
(a) Renderings of the captured geometry
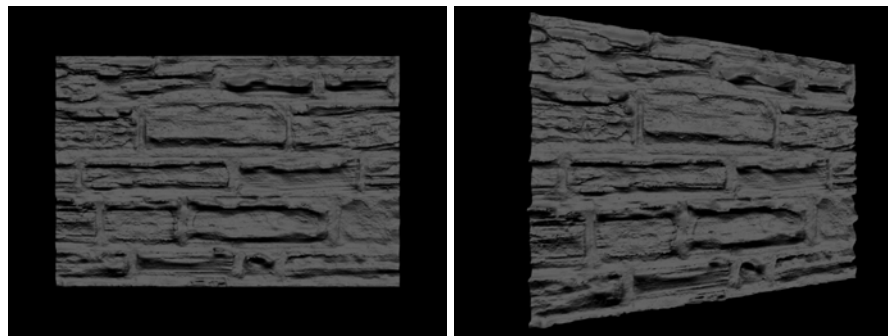
(b) Renderings of the transfered geometry using HM
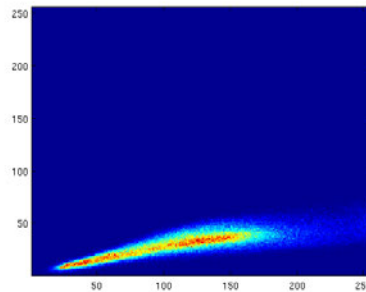
(c) Two dimensional histogram

Figure 3.16: *Synthetic renders of the captured and transferred meso-structure for exemplar $ChapBrick8$ and its H^2 histogram.*

the albedo producing visually rich models due to the high frequency detail. Strong geometric features like small protrusions are correctly estimated since their appearance under ambient lighting is more correlated with depth and therefore the transfer is correct. This demonstrates that orientation in high frequency geometry is well approximated even in difficult cases like this one. In large areas where colour in the albedo is dark or bright, and the shading does not correspond to indentations and protrusions respectively, large artifacts appear.

The exemplar $Brick1$ has a highly spread shading-ambient 2D histogram. For comparison, we show in figure 3.16 the results of another exemplar where the 2D histogram is closer to a monotonic curve with a narrower distribution so the transfer is predicted to work better. We observe that the final meso-structure is practically indistinguishable from the captured exemplar.

Under the assumption that the same material has similar shading-appearance distributions, we can predict the behaviour of the transferring by looking at its H^2.

Exemplar Brick1: Albedo Artifacts

(a) Albedo maps: From left to right, captured, histogram matched, difference image

(b) Model rendered using captured albedo

(c) Model rendered using histogram matched albedo

Figure 3.17: *Albedo artifacts are reduced in the final renderings because of shading consistency.*

The transferred albedo map for the exemplar $Brick1$ presents smaller differences than the shading observed in Figure 3.17. Comparing the captured and the transferred albedo maps, some shading remains in the histogram matched albedo and the higher errors are present in shaded areas as predicted by the theoretical analysis. The importance of these artifacts in the shaded areas is reduced when producing the final

65

renderings since these areas are very likely to be shaded by the indentations in the meso-structure as they were in the original image. Figures 3.17(b) and 3.17(c) show two renderings under novel illumination using the same geometry textured with the recovered and histogram matched albedo respectively. The histogram-matched albedo rendering presents some visible overshading, but this is less noticeable than when comparing the albedo maps from Figure 3.17(a). Computing the difference image of the renderings, we get a mean error of 6.8% while the difference in the albedo map was 10.5%. We tested this hypothesis using different lighting conditions and results always present an improvement regarding the differences in the albedo maps.

3.3.4 Conclusion on Histogram Matching

We have analysed the performance of Histogram Matching as a material transfer technique and evaluated it numerically to complement the previous perceptual study included in the depth hallucination paper [GWJ+08]. In the results of that study, the examples using histogram matched models, were rated marginally lower than the captured ones, but still achieved high ratings as real images. In our analysis we conclude that the performance depends on the shape of the data distribution of the ambient map and the corresponding channel we want to transfer. The closer this distribution is to a monotonic curve the better the accuracy of the transfer. We also conclude that artifacts in the albedo map are partially masked in the final rendering because of its correlation with the shading areas. Finally, we observe that low frequency disparities in the shading map produce more noticeable artifacts than high frequency ones.

In summary, histogram matching is able to successfully transfer albedo and meso-structure with average errors of 7% for albedo maps and 10% for shading maps and has been shown in trials with human subjects to yield perceptually plausible results.

3.4 Transfer by Analogy

The second transfer method is also inspired by the texture synthesis literature. This time, instead of a stochastic method, we propose an image-based algorithm. Over the last decade texture synthesis by exemplar has been a very active area of research. We refer the reader to recent state of the art reports for a complete review [KW07, WLKT09]. Using exemplars for synthesising texture has been proved to work effectively for synthesising a large variety of textures. New pixels or patches are generated

by choosing the best candidate from a given exemplar, such that it is coherent with the already synthesised texture. For globally-varying textures, a control map is often used to drive this type of synthesis. Ashikhmin [Ash01] synthesised an output conditioned by a user drawn coloured map, and similar ideas are used in patch-based synthesis [EF01] using an image as a guide. An extension of this was the notion of *image analogies* [HJO+01]. This method takes three images as input: a reference exemplar, a filtered version, and a target image, and transfers this filter to the target image, creating its analogous filtered image.
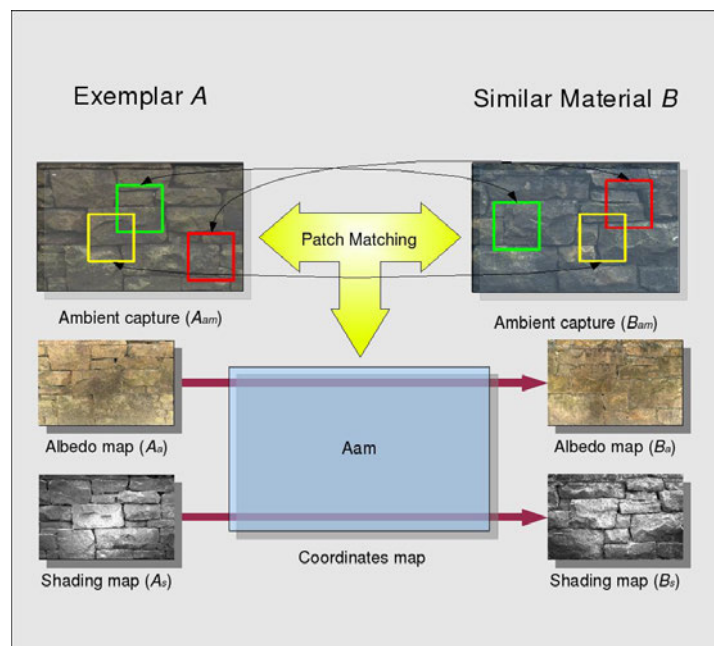
Figure 3.18: *Transfer by Analogy Process.*

We apply the concept of *image analogies* to our problem. We consider the albedo and shading images as filters of the ambient map. The algorithm is able to reproduce these filters for a new image, in our case the new sample of the same material, producing a new corresponding albedo and shading image. The basic idea is illustrated in Figure 3.18.

By comparing A_{am} and B_{am}, we find the best match in A_{am} for every pixel in B_{am}. This provides a coordinate map associating every pixel in B_{am} with a pixel in the exemplar A. Using the coordinate map and the filtered source images (A_a,A_s), we create the new filtered target images (B_a and B_s). The algorithm is as shown in Figure 3.19.

The core of this algorithm resides in finding the *best match* within A_{am} for every

```
FUNCTION TransferByAnalogy(A_am, A_s, A_a, B_am, &B_s, &B_a)
{
   Compute_Gaussian_Pyramids(A_am);
   Compute_Gaussian_Pyramids(B_am);
   Compute_Features(A_am);
   Compute_Features(B_am);
   FOR EACH (level L)// from coarsest to finest
   {
      FOR EACH (pixel q) IN B_am
      {
         p <- BESTMATCH(A_am , B_am , q , L);
         coordinatesMap(q) <- p;
      }
   }
   reconstructFromCoordinates(coordinatesMap, B_a) ;
   reconstructFromCoordinates(coordinatesMap, B_s) ;
}
```

Figure 3.19: *Transfer by Analogy Algorithm.*

pixel in B_{am}. A best match is defined as the minimum distance in a feature space so an appropriate feature vector in this space must be defined for every pixel. This feature space needs to be able to transfer the relationship between the filtered and the unfiltered images.

3.4.1 Feature Descriptor

A feature descriptor for a pixel is an array, whose length n defines its number of dimensions, containing a series of values that can be computed for this pixel and defining a vector in n dimensional space. The first feature descriptor that we can think of for a pixel in the ambient image is the vector containing its colour triplet. As we have seen in section 3.2.1, the relationship between the ambient image and the shading and albedo maps describes a distribution that depends on the material. For every value in the ambient image there is a distribution of possible values in the other two maps. Using the colour triplet means that for every pixel there will be many pixels with the same feature descriptor but with different values for the filtered images (albedo and shading).

Finding the best match using this feature space leads to different solutions, but not all of them are necessarily right. We need therefore to establish a feature space where this relationship is able to determine a good albedo and shading value for a given pixel. We would ideally want a feature descriptor that has a deterministic relationship with the filter (albedo and shading) so the solution is unique and valid. This feature descriptor has to be able to be computed from the information contained in the Ambient map

alone, since this is the only information available for the maps to compare A_{am} and
B_{am}. The feature descriptor has to be transferable, i.e. has to be able to computed in
two different images and produce the same feature vector. Also, the distance within
this feature space must represent how well a vector descriptor transfers the filters.

In the texture transfer and exemplar based texture synthesis literature, researchers
have used neighbourhood information to keep coherence in the synthesised image [WL00a,
HJO$^+$01]. Use of these extra channels like feature masks [WY04, ZZV$^+$03] and dis-
tance to feature [LH06] have been proposed to better retain global structure in semi-
regular textures during synthesis. These feature masks are computed manually, so we
would need to create a feature mask for every exemplar and this has to be reproduced
in the image to be matched. To compute these feature masks automatically is still an
open issue.

Using neighbourhood information helps to constrain the matching, creating a more
deterministic feature vector. We use 7x7 pixels patches, as in [HJO$^+$01], but we use
the three RGB channels instead of just the luminance channel Y. We can afford this
extra computation due to improvements we made to the algorithm to find the best corre-
sponding pixel, as described in section 3.4.2. For every pixel we produce a 7x7x3 (147)
high-dimensional feature vector. For a given pixel value in B_{am}, the pixel from A_{am}
that best matches it is the one with the same values for the given filtered images.

Since there is significant variation from one exemplar to another, it very difficult
to define a general feature descriptor that works optimally for all possible materials.
We studied the relationship between the different channels of the ambient map, the
gradient maps, second gradient, and luminance channel with the filters. We also tested
different patch sizes from 1 to 9 (we need to limit the size to 9 because of memory
limitations). We compute the mutual information between the feature descriptors and
the corresponding patch in the filter. The results show variation between channels,
with the green component being the one that shares most information. However, this
changed from exemplar to exemplar and therefore, the results could not be generalised.
Since variability in the input data did not provide evidence in our experiment, we
drew inspiration from recent texture synthesis algorithms [LH06] and fixed our feature
descriptors based on patches around the pixel, where 7x7 RGB patches have been
proved to perform correctly for a variety of textures. We evaluated the algorithm using
this descriptor, resulting in similar transfer accuracy to Histogram matching. However
we found that the nature of the artifacts is different.

Empirically, we have tested several feature descriptors, such as different colour

spaces, different ways of computing masks, patch sizes, etc. Unfortunately since there
is significant variation from one exemplar to another, it very difficult to define a general
feature descriptor that works optimally for all possible materials. The one described
here provides plausible results for many cases, but ideally an adaptive descriptor would
be desirable. Recent research carried out by Eric Risser at al. [RHDG10] found that
sampling textures at different scales simultaneously keeps global and local structure
in texture synthesis. Our algorithm is general and can take advantage of these ad-
vances in the development of feature descriptors for textures. For the particular prob-
lem of albedo and meso-structure, we believe that future work in feature descriptors is
a promising area of research.

3.4.2 Nearest Neighbour Search

Finding the best match is a nearest neighbour search. Due to the large search space
(millions of pixels), the algorithm must be efficient. Nearest neighbouring search for
patches in a medium/large sized image rapidly becomes prohibitively expensive. Since
patch-based sampling methods have become popular for image and video synthesis, re-
searchers have studied optimising this process [WL00b, WSI07, KFCO$^+$07, KZN08].
We employ a recent fast patch matching algorithm for approximate nearest-neighbour
matching between image patches [BSFG09] that performs at interactive rates in their
implementation.

The algorithm begins with a random initialisation, and then uses an iterative pro-
cess consisting of two steps: a propagation that searches within the neighbourhood
of the previously matched pixel, and a random search in the further area that helps to
avoid a local minimum. This iterative process proceeds in scanline order (from left to
right, top to bottom) for odd iterations, and in the inverse order for even iterations. The
final Transfer by Analogy algorithm becomes as shown in Figure 3.20.

The algorithm converges quickly giving a good level of detail in 5 iterations. The
key of this algorithm is the assumption that neighbouring source patches are likely to
be matched by neighbouring target patches which is denoted as *coherence*. This finding
was already explored in exemplar based synthesis algorithms to restrict the search for
ideal new patches [Ash01]. The second important aspect is random initialisation. Due
to the big number of samples, the probability of one of these random samples resulting
in a good match is very high. Once one good sample is found, the propagation phase
propagates the good matches. Running this algorithm in a coarse to fine fashion and
including the random search, makes it converge to a good global solution very quickly,

```
FUNCTION TransferByAnalogy(A_am, A_s, A_a, B_am, &B_s, &B_a)
{
  Compute_Gaussian_Pyramids(A_am);
  Compute_Gaussian_Pyramids(B_am);
  Compute_Features(A_am);
  Compute_Features(B_am);
  FOR EACH (level L)// from coarsest to finest
  {
    IF (L == coarsestLevel)
        InitialisazeRandom(coordinatesMap);
    FOR iteration IN Number_of_Iterations
        {
            FOR EACH (pixel(q,iteration)) IN B_am //pixel proceeds in scaline order or inverse order depending on the iteration
            {
                p_old <- coordinatesMap(q);
                p_new1 <- Propagation(q,coordinatesMap);
                p_new2 <- RandomSearch(q,coordinatesMap);
                p <- best_of(p_old,p_new1,pnew_2);
                coordinatesMap(q) <- p;
            }
        }
    escaleUP(coordinatesMap);
  }
  reconstructFromCoordinates(coordinatesMap, B_a)  ;
  reconstructFromCoordinates(coordinatesMap, B_s)  ;
}
```

Figure 3.20: *Fast Transfer by Analogy Algorithm.*

which makes it possible to use in our high resolution images in a reasonable time (30
seconds for a 700 image resolution, 7×7 patch size, 5 iteration, and 5 levels of scale).

3.4.3   Image Reconstruction from Patches

The search algorithm returns a coordinate map containing the best match for every
pixel. Since our matching is performed in patches, every pixel is contained in as many
overlapping patches as theres are pixels within a patch. Several strategies can be de-
vised to reconstruct the final image from the coordinates of the best-match map. The
first trivial one is to directly take the value of the pixel in the coordinates determined
by this map. Second is to average equally all the pixels. The last one is *pixel vot-
ing* [SCSI08]. Pixel voting is a weighted average according to the distance to the
accuracy of the match in this pixel for a given patch.

Figure 3.21 shows the results of the three different strategies. Direct reconstruction
produces hard transitions between colours and makes it look artificial. Equal averaging
makes transitions softer but produces over-blurring.

Pixel voting produces smooth transitions between patches and sharper results than
simple averaging. Since it is computed per pixel, it allows us to correct for partial
matching. For instance, if the best matched patch has a pixel that is very different,
the particular quality of the matching for this pixel is taken into account, getting a low

71

(a) Direct Reconstruction

(b) Equal Average

(c) Pixel Voting

Figure 3.21: *Results of the three different strategies for image reconstruction from patch coordinates*

voting value for this specific pixel when creating the final image.

3.4.4 Experimental Evaluation

The evaluation of the method *Transfer by Analogy* needs a different experiment to
the *Histogram Matching* one. In this case, if we use the same ambient image for
A_{am} and B_{am} the filtered images resulting the transfer will be identical to the original
maps. Instead, we need to define pairs of exemplar of the same material. We take the
same exemplars used previously and define pairs and triplets where the material is the
same. Then we run the transfer between them. For example, for the same material
redbrick, we have two exemplars $BRICK1$ and $BRICK3$. We use $BRICK3$ as
an exemplar to infer the albedo and shading map of $BRICK1$. Since we have the
complete exemplar for $BRICK1$, we can compare the results of the transfered maps
with the captured ones. We can also run the symmetric experiment, using $BRICK1$ as
exemplar to infer the material properties of $BRICK3$. Therefore we have two results
for every pair, and six for a triplet.
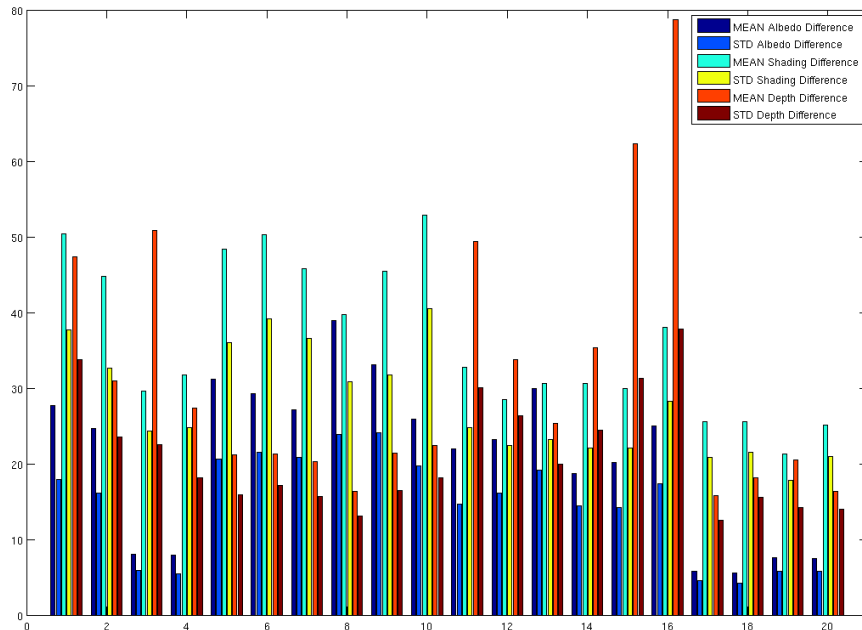
Figure 3.22: *Results Transfer by Analogy.*

We computed the differences with the reference exemplars. Figure 3.22 plots the
mean and the standard distribution of the difference as in section 3.3.2. The difference
means are paired by material. In general, if an exemplar A is a good match for B, B is
a good match for A. Globally we see that albedo was better transferred than shading.

It is interesting to note that in some cases (2-3, 17-18, 19-20) the albedo transfer was
really accurate, with a error mean under 3%. Most (85%) of the exemplar presented
a mean difference in albedo around 10%. Shading presented higher errors, between
10% and 20%. The depth estimator reduces the differences of the shading maps in
most cases as explained for the histogram matching evaluation. However, as with
histogram matching, in certain cases the depth difference shows a greater difference
than the shading (see 3,11,12,14,15,16). The explanation for this phenomenon is the
same as previously. If a large region in the shading map is transferred incorrectly, that
produces a large disparity when comparing the resulting depth maps.

The nature of the error in this transfer method is more difficult to predict. The
matching will depend on whether the features in the reference image are present in the
exemplar. The matching is local so so it is expected to work better with more regular
materials. We confirm that in Exemplar pairs 17-18 and 19-20, where the material is
closer to regular and there is even some overlapping between the captures. The transfer
in this case, does not depend on the global appearance of the albedo or shading map,
but on the appearance of the material under diffuse lighting which was ignored by the
histogram matching method. However, when the features are not present, the resulting
filtered images are likely to introduce structural artifacts. Also the final result lacks
high frequencies due to the reconstruction process. This is particularly important in
the shading image, since as we explained before, the high frequencies are important
perceptually for humans to estimate the meso-structure. The quality of the transfer
will depend on the actual correspondence of the relationship appearance-shading and
appearance-albedo in both exemplars.

Figure 3.23 shows some results of transferred albedo maps. We observed that the
results vary in quality depending on the exemplar we use for transfer. When the exem-
plars match each other appropriately (Figures 3.23(b), 3.23(d), 3.23(e)), the resulting
albedo is very similar to the original one and does not suffer from over shading arti-
facts. On the other hand, this method is more sensitive to have a good match. When
exemplars do not correspond very well (Figures 3.23(a), 3.23(c)) the results lose struc-
ture and coherence. This method can produce very accurate results but it is more
sensitive to exemplars matching the materials appropriately.

(a) Left to right(A,B,C,D); A: Albedo Exemplar X; B: Transferred Albedo for Exemplar X using C; C: Albedo Exemplar Y; D: B: Transferred Albedo for Exemplar Y using A
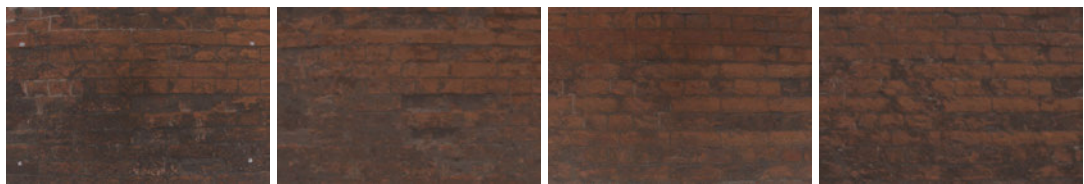
(b)

(c)

(d)
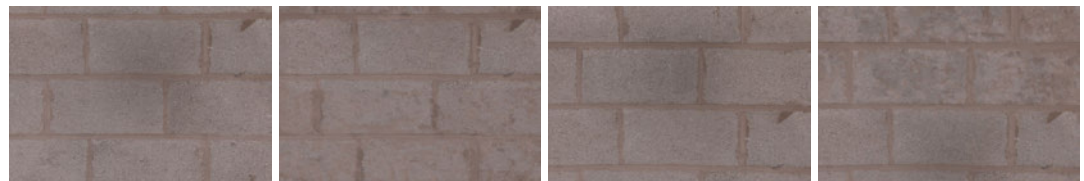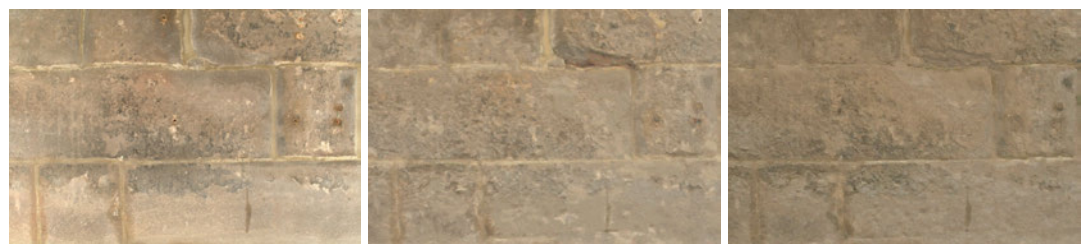
(e) From left to right: Albedo Exemplar X, Transferred Albedo for Exemplar X using Exemplar Y; Transferred Albedo for Exemplar X using Exemplar Z

Figure 3.23: *Four albedo pairs and a triplet results of the transfer.*

3.5 Discussion and Comparison of Methods

We compare the quality of the transfer of our two methods numerically. The first stage is to run the same evaluation experiment. We take the same exemplar pairs and triplets

used to evaluate the Transfer by Analogy technique and run the Histogram Matching
transfer on them. The results are shown in figure 3.24.
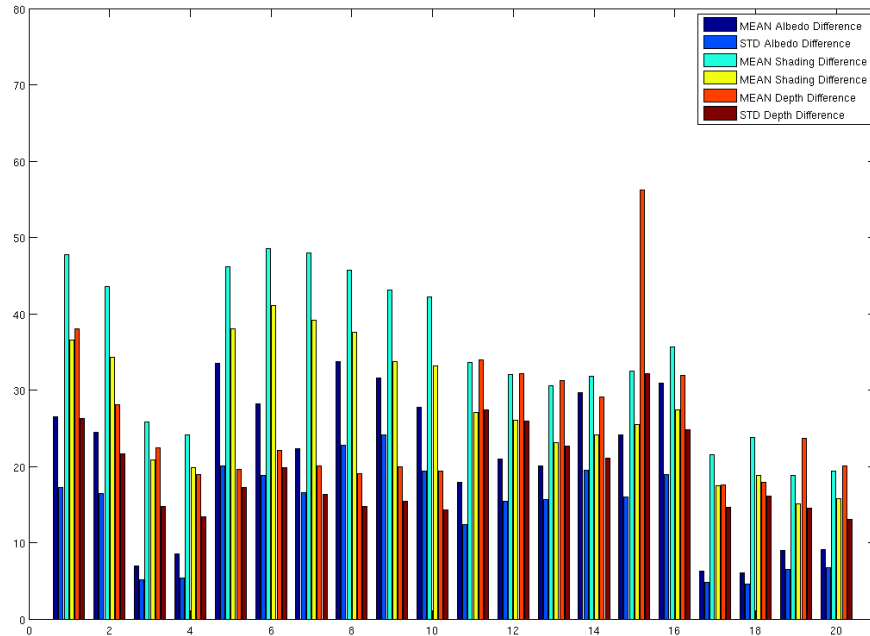
Figure 3.24: *Results Transfer by Histogram Matching.*

We compared side by side the resulting statistics of both methods in the transfer
of albedo, shading, and the resulting depth from the shading map (figure 3.25). The
results show that both albedo and shading are transfered with similar accuracy, i.e. the
final images produced are numerically equivalent in terms of error. However the final
depth is in general somewhat better in the case of Histogram Matching. Depth maps
synthesised using Transfer by Analogy lost high frequencies in the process and con-
sequently the structural coherence is sometimes missed. Histogram matching created
more coherent shading images that lead to more plausible depth maps.

The conclusion that we extract from this comparison is that the general accuracy is
similar. However, the nature of the errors introduced are different. Histogram match-
ing is more sensitive to the colour-depth discrepancy having problems to fully remove
shadows and may misinterpret dark albedo as indentations and bright albedo as protru-
sions. However, it preserves very well the global structure of the texture, and produces
plausible results. On the other hand, Transfer by Analogy is less sensitive to the global
statistics because of the local matching. This is an important quality because allows
us to match materials where the global statics change, which is common when trans-
ferring to a larger surface. We will discuss this in detail in chapter 4. This method
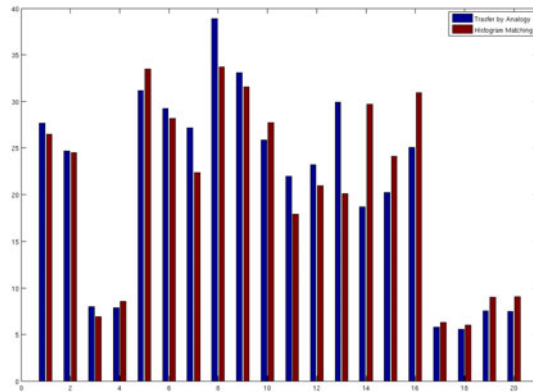
(a) Comparison albedo maps
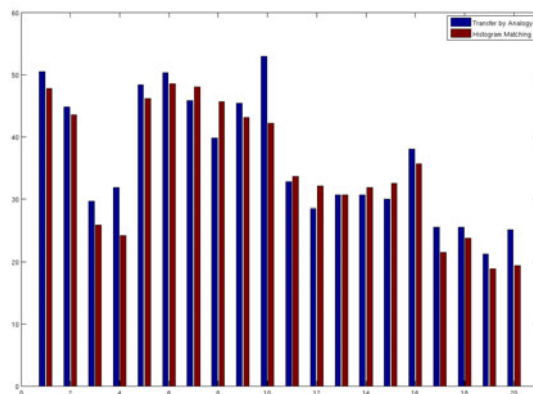
(b) Comparison Shading
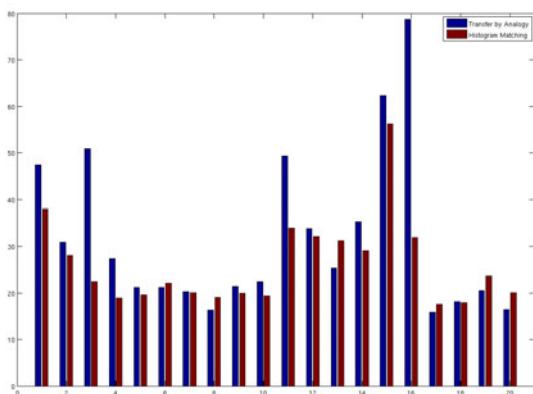
(c) Comparison Depth

Figure 3.25: *Error comparison between two transfer methods: Transfer by Analogy (Blue) and Histogram Matching (Red).*

removes shadows effectively by solving the albedo-shading ambiguity locally instead of globally. In the negative aspect, the final result can lack high frequency detail in

comparison with histogram matching, which is particularly important for the plausibility of the meso-structure.

This new method provides a number of important advantages over histogram matching. By performing local matching, the technique is less sensitive to the quality of the match between the global statistics of the exemplar and the target image. This is important since exemplars are in general small samples of the material and might contain features that are statistically significant in the exemplar but not in the global texture, or vice versa. Another important characteristic is that we can search for the best patch in several exemplars side-stepping the need for segmenting and fixing a material for every segment. This is studied in chapter 4. Also, since we combine the different patches according to the quality of the match using pixel voting, we can reproduce features that are not directly present in any one single exemplar by mixing several of them.

Finally, since the matching process produces a coordinate association between the exemplar and the target image, we can potentially increase the resolution [HJO⁺01] of the textures when the exemplar has a higher resolution than the global texture. Note that this is normally the case, since exemplars are typically captured as close up-views.

Transfer by Analogy behaves better with patterned textures since they are close to regular textures and stochastic textures are better transferred by Histogram Matching.

3.6 Conclusions and Future work

This chapter has presented a method to capture reflectance and meso-structure sample for globally flat textured material, which is able to create perceptually plausible renderings of these surfaces. It presents, analyses, and evaluates two image based methods to transfer both reflectance maps and shading maps: Histogram Matching and Transfer by Analogy. The transfer capabilities and limitations are studied for both methods, comparing and discussing the qualitative evaluation as well as the nature of the artifacts produced by them.

We reviewed an inexpensive albedo and meso-structure capturing technique named Surface Depth Hallucination. We use this technique to capture exemplars of textured materials, then we define the problem of un-lighting textures and inferring geometric detail as a filter transfer problem based on such exemplars. The aim of proposing transfer techniques to infer albedo and meso-structure is to overcome the limitation in capturing these properties imposed by the necessity of perform it in controlled lighting environments. We analysed the behaviour of two different transfer methods: histogram

matching and transfer by analogy. The relative merits of these two approaches were discussed. We concluded that both techniques reasonably accurately transfer both material properties, although the performance of both techniques is data dependant. Histogram Matching depends on the relationships between the material appearance and its properties, and provides explanations for meso-structure that are plausible and structurally consistent with the appearance. Transfer by Analogy depends on the similarity between the exemplar and the target material.

As future work, we propose investigating transfer techniques that combine both proposed methods taking advantage of their benefits. As stated in the previous section, the nature of the artifacts as well as their "strong points" are different and up to a point are complementary. In the case of albedo, Histogram matching fails to fully remove shadows but generates highly detailed maps; Transfer by Analogy results in a loss of high frequency detail, but shadows are correctly eliminated. In Figure 3.26, we show a simple preliminary experiment of a possible combination, by merging both albedo maps resulting from both transfer techniques. The way these two images are combined is by taking the brightest pixels of both of them. That effectively removes the remaining shadows in the histogram matched albedo and, at the same time, keeps most of its detail. More complex combination techniques and new transfer options are left for future work. Also, further studies on the relationship between reflectance, shading, and appearance could lead to feature descriptors with better transfer characteristics.

(a) Histogram Matched Albedo. (b) Transferred by Analogy Albedo. (c) Transferred Albedo using our hybrid method.

Figure 3.26: *Preliminary result for future hybrid transfer method. This method removes the shadows effectively and preserves high frequencies.*

The ultimate aim of the transfer technique is to recover the necessary properties of the materials for building façades. A complete façade normally contains different materials in different proportions. Chapter 4 studies how to extend the transfer techniques presented here, to construct complete façades.

Transferring Material Properties to Large Surfaces

THIS *chapter deals with the problem of transferring material properties to an image where several materials cover large surfaces. We present a solution to deal with several materials for both transfer techniques* Histogram Matching *and* Transfer by Analogy. *We also investigate an extension to the* Transfer by Analogy *algorithm to automatically associate the different exemplars with the different materials in building façades.*

4.1 Introduction

In chapter 3, we have introduced two techniques to transfer albedo and meso-structure from exemplars: Histogram Matching and Transfer by Analogy. We evaluated these techniques using pairs of exemplars of the same material. The real advantage of a transfer approach is that it allows us to approximate these properties for large surfaces, reducing the capture effort and inferring these material properties for inaccessible sections where capture is infeasible. Photographs of larger surfaces, and particularly of building façades, typically contain different materials. Both our transfer methods work under the condition that the exemplar and target image contain the same single material. Therefore, they cannot be applied directly in this scenario. This chapter studies the problems involved in the transfer of albedo and meso-structure to large surfaces such as building façades.

Our solution consists of separating the different regions in the image, according to a homogeneity predicate based on a material model, and providing associations between these regions and their corresponding exemplars. Our separation and association problem can be stated as: given a façade with several materials and a set of exemplars containing all the materials present in the façade, we aim to identify regions that have

the same underlying material properties and can be associated in a one-to-one relationship with the exemplars.

The homogeneity predicate required for this process is defined by the characteristics of the different materials, but also depends on the transfer method used. A material is a textured surface that contains the same albedo properties and shading properties as the texture captured in an exemplar. In the case of Histogram Matching, we require material properties to be globally and statistically the same. Specifically, exemplar and segment must share albedo and shading histograms. In the case of Transfer by Analogy (TbA), the same material can be defined as the one whose local features are contained in the exemplar. Having two different requirements for the homogeneity predicate means that the desired segmentations for the different transfer techniques might differ from each other. At this point in our exposition, we assume that exemplars contain a unique material and both global and local homogeneity predicates are satisfied.

The process of separating regions of an image, according to homogeneity predicate, is usually referred as segmentation. Image segmentation is a complex, application-dependent problem. Applications include background/foreground separation, object counting and recognition, image retrieval, scene interpretation, and texture separation. It is important to state that we do not try to solve the image segmentation problem which is more general, but finding an appropriate solution for our specific problem by exploring ways to efficiently transfer material properties to façades with several types of textures. In this chapter, we explore the application of segmentation techniques to our problem. We present a system developed based on *Graph-cuts* [BVZ01] to interactively segment the material in an image. Furthermore, aiming at a completely automatic process, we investigate the performance of data-driven texture descriptors for automatic material association, by extending our Transfer by Analogy method to account for several materials and exemplars. The results of this investigation are promising and encourage further research on this technique.

4.2 Graph-cut Based Semi-automatic Material Segmentation

First, we developed an interactive system for material segmentation. Interactive techniques allow for robust segmentation giving the user the opportunity to correct errors in the automatic process. Also, user input provides important guidance for texture separation and association, taking advantage of the human ability to differentiate materials

and textures.

The user draws strokes over the image, marking areas of different materials and associating a label with each material. Then the algorithm optimises the segmentation according to the user input and the image information. The user can refine the segmentation by drawing more strokes and re-optimise until the result is satisfactory.

The optimisation process uses *Graph-cuts* [BVZ01] which finds a globally optimal division of the image, finding a label per pixel that minimises an energy function. The image is represented as a graph where every pixel is a node connected to its four neighbours by edges. Every possible material is marked by a stroke, and assigned to a label. An energy function is defined for each node and edge, then graph-cuts are used to find an approximate solution to the minimal energy configuration. The energy function to be minimised contains two sub-functions defined by the following equation

$$E(X) = \lambda \sum_{i} E_{data}(x_i) + (1 - \lambda) \sum_{i,j} E_{smooth}(x_i, x_j) \qquad (4.1)$$

where E_{data} defines the data cost at every node x_i and specifies a cost of assigning a label to this node. This reflects the likelihood that is this node belongs to this label. E_{smooth} is the smoothness term and defines the cost of assigning different labels to two neighbouring nodes. Finally, λ establishes the relative importance of one sub-function against the other. A high λ value prioritises the data cost minimisation, allowing for a bigger number of cuts, and therefore smaller and more numerous segments. A low λ value gives more importance to the smoothness term, encouraging large continuous segments.

The user marks-up each different material in the texture map by drawing a stroke over the regions composed of the same material, as shown in Figure 4.1. A different coloured stroke is used for each material. The pixels in the image covered by the strokes of each colour are used to create an appearance model of this material. The RGB values of these pixels are clustered using K-means, and the centroids of these clusters define the appearance model of this material. We set the algorithm to have 64 clusters as in [XFT+06], which gives enough colour variety to the appearance model to account for colour variations. The cost of assigning a node in the image to a material is the minimal euclidean distance (in the RGB colour space) between the node colour and the centroids of the appearance models, a short distance meaning a low cost. Following equation 4.2, every node is assigned a cost for being labelled with a material.

(a) Strokes   (b) Segmentation
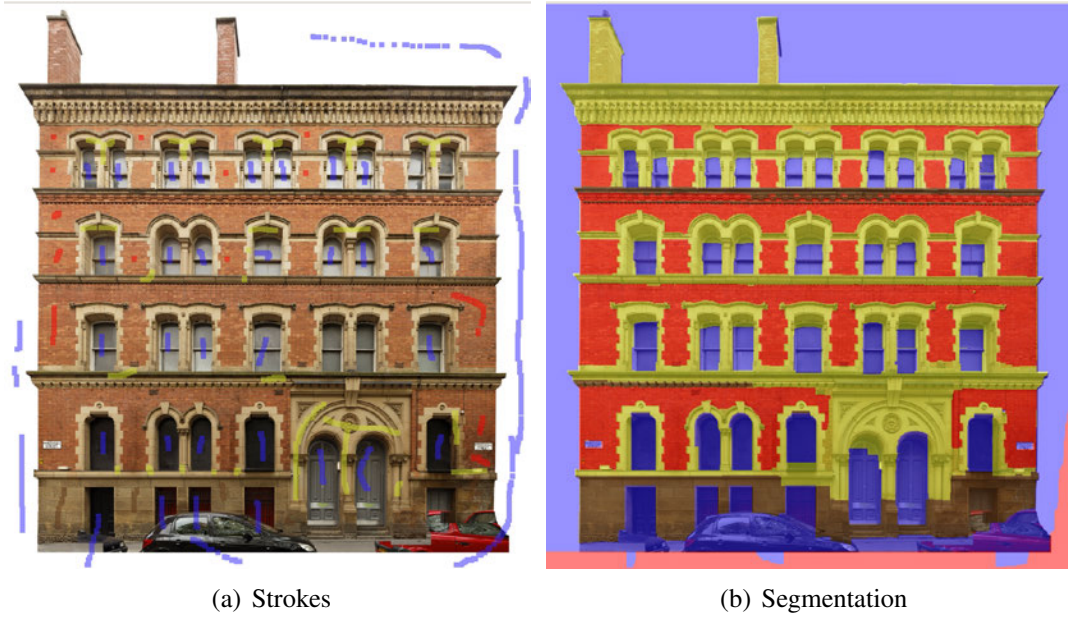
Figure 4.1: *Strokes and resulting segmentation using our interactive system*

$$E_{data}(x_i) = min\_distance(Colour(x_i), Centroid_j) \forall j \in \{1..64\} \qquad (4.2)$$

The smoothness term is defined for every edge connecting two nodes in the graph, setting a cost for assigning different labels to neighbouring nodes. This cost is a function of the colour difference (in a Euclidean distance sense) between adjacent nodes. The bigger the difference the lower the cost, encouraging cuts in the high-contrast edges and continuity in low-contrast areas. The smoothness function E_{smooth} can be written as

$$E_{smooth}(x_i, x_j) = distance(x_i, x_j)^{-1} e^{-\beta(x_i - x_j)^2} \qquad (4.3)$$

where $\beta = (2 \langle (x_i - x_j)^2 \rangle)^{-1}$ and $\langle \cdot \rangle$ denotes expectation over an image sample [RKB04].

Once the costs have been established, the labelling problem is solved by minimising the energy function over the whole graph. Several algorithms can be used to find an approximation to a global solution [SZS+08a]. We use graph-cuts to solve the optimisation, which provides a good balance between performance and execution time.

The initial segmentation can be refined interactively by drawing more strokes, recalculating the appearance model and the data cost. Once a satisfactory segmentation is obtained, the user links exemplars and materials through a simple user interface.

We demonstrate our segmentation system with a façade test case in Figure 4.1. The segmentation process is quick and intuitive. The strokes in Figure 4.1(a) are the result of two refining iterations. Each iteration took approximately one minute for the user to draw the strokes and approximately two minutes for the algorithm to find a close-to-optimal solution to the segmentation. In this way, the association segment-exemplar is performed manually. Since there are just a few materials and a small number of segments, this task is both simple and convenient. In a total time of less than 5 minutes, the user obtains a good segmentation solution that can be refined to the desired level of accuracy. In contrast, to segment this image (note that its resolution is 2300×1800) using a completely manual process, marking the boundaries of the regions with a line drawing tool, took around 2 hours of work. Our semi-automatic method presents therefore an important time reduction compared to completely manual segmentation, proving to be an adequate tool for a typical façade segmentation.

This approach works adequately when boundaries between materials are well defined as in this example. However, problems arise in places where the architectural features are complicated. Gradient and colour based segmentation techniques, such as ours, tend to locate segment boundaries in high-gradient areas. When the desired boundary should not lie in these areas, the algorithm may behave poorly. User guided refinements normally solve these problems at the cost of more iterations.

It is important to keep in mind some guidelines when capturing exemplars in order to simplify the process and get optimal results. For instance, an exemplar containing both a brick wall and moss, requires segments to contain proportional regions of brick and moss for good transfer. This complicates the segmentation process. Therefore, it is normally a good practice to capture exemplars that contain a single material. We can remove the area with moss from the exemplar, or divide the exemplar in two: moss and brick wall, making the segmentation more intuitive and easier to perform. On the other hand, middle-size segments generally match the material statistics better than large segments, since the latter are more likely to contain irregularities.

Using our system we are able to create material associations and segments for our transfer methods in a semi-automatic way, reducing the time and effort spent in this stage of the system. Interaction is simple and intuitive, but the process relies on the judgement of the user and, in very challenging cases, the amount of refinement required can lead to times similar to those for fully manual segmentation. New segmentation algorithms are constantly being presented, improving previous approaches. Due to time constraints, these improvements have not been deeply explored, but we suggest

this research for future work.

As an alternative to complex segmentation techniques, we explored the extension of our TbA algorithm to several materials. This has three interesting goals. First, if our matching metric is capable of correctly matching the appropriate material from a set of given ones, explicit segmentation is not necessary, since it is implicit in the transfer technique. Second, this would lead to a completely automatic segmentation-association-transfer process. Third, if the association provided by the algorithm is correct, the feature space, matching metric, and algorithm, would be interesting for texture segmentation. The next section presents our novel extension of the TbA method to several materials and presents the results of the automatic transfer of material properties to a full façade.

4.3 A Novel Approach to Automatically Associate Texture and Exemplars

Interactive or manual segmentation is a working solution for transferring several materials in an image, demonstrating particularly good behaviour when the boundaries between materials are well defined. When these boundaries are not clear, segmentation can be difficult. Therefore, it is interesting to investigate ways to avoid the segmentation process. The TbA algorithm associates every pixel in the target image with its best match within an exemplar. Extending this algorithm to process several materials chooses the best match within a set of exemplars without the necessity of previously segmenting the image. This automatically provides a per-pixel association with the material, which provides a segmentation, but with an important conceptual difference. Segmentation encourages continuity of segments, whereas our novel association encourages the best local match. This difference has a number of interesting consequences:

- It separates the concept of exemplar from the concept of material; exemplars may contain several different textures or have local variations (such as moss in a brick wall), and the algorithm would still detect the best match. That simplifies the exemplar capture process.

- It is a general algorithm and completely automatic, so it has the potential to be extended to n exemplars in a database. This would make it unnecessary to capture

the exemplars in-situ: and would allow us to perform the transfers to already existing models.

- It allows capturing several exemplars of the same texture to collect a larger set of local features, leading to a better transfer.

- It allows for partial matching. That means that if a certain small region within a texture has a better match to a different exemplar, we allow this possibility and encourage the best match within all the exemplars.

- We can perform a "fuzzy segmentation" – assigning weights to exemplars instead fixed labels – which allows reconstruction of textures that are not present in a single exemplar but are a combination of several of them.

- The advantages of having an association between every pixel of the texture map and a higher resolution exemplar allows us to increase the resolution of the texture map.

- The segmentation is actually defined by the exemplar. Each exemplar acts as a complete very rich vocabulary of features instead of a fixed texture model.

- Finally, this approach allows fine grain segmentation, which would be very time consuming to do manually.

These potential advantages encouraged us to investigate such an extension of the TbA algorithm.

Several aspects of the technique need to be investigated. In chapter 3 we tested the ability of the algorithm to match features in order to transfer material properties. In this chapter, we evaluate the algorithm and the matching metric when applied to several materials.

We experiment with data-driven texture descriptors for our transfer and association technique. Specifically, our extended TbA algorithm uses small 7×7 pixel patches. These pixel features are inspired by data-driven texture synthesis techniques [EF01, Ash01, Ash03] which have proven to be able to produce very realistic textures from small exemplars by using patches and neighbouring data. These methods have been evaluated by users that perceive the synthetic results to be the same texture as the exemplars. This shows that most of the information important to reconstruct texture is contained in the close neighbourhood. When applying these techniques in a multi-scale approach, the global structure of the texture is well preserved during synthesis. A recent approach [RHDG10] uses the same type of texture descriptor but

samples several scales at the same time, which deals with the variation across scale. We discuss the problem of scale in more detail in section 4.3.6. Our observation is coherent with the texture analysis and segmentation literature where descriptors based on small patches like textons (minimal texture entities normally defined by a small patch) [ZGWX05], Gabor filters [JF91] or texels (distinct cohesive groups of spatially repeating patterns) [TA09] are used to extract statistical models of texture. These and other texture features have not been explored within this thesis for time constraints. Further studies on texture models based on more complex descriptors could benefit our algorithm improving the matching process.

In the rest of this chapter we introduce our suggested extension and evaluate our matching metric with synthetic and real experiments. We discuss the ability of the algorithm to associate the correct materials and exemplars, and the robustness of the association when increasing the number of possible labels (exemplars). It is worth emphasising again that we are not trying to solve the texture segmentation problem, but to investigate the utility of data driven texture descriptors to automatically transfer and associate materials, which results in a segmentation, but does not have the same requirements as traditional segmentation.

4.3.1   Description of the Method

The TbA method finds the closest match between every pixel in the texture map and an exemplar, based on the assumption that two pixels with a similar appearance (described by a patch of neighbouring pixels) will have similar underlying material properties. We propose to extend this method to several materials, which leads to a novel exemplar-based data-driven technique that performs segmentation, association, and transfer simultaneously and automatically. We rely on a similar assumption we made for the TbA approach in chapter 3: the same material (in an exemplar and in a photo-texture) under sufficiently similar diffuse lighting conditions can be correctly associated by comparing their appearance.

In order to extend the algorithm to handle several materials, we find the best per-pixel association for every exemplar separately and store its coordinates. This process gives us an error metric for every pixel and its best match in every exemplar. In our results we labeled pixels with the exemplar that presents the smallest error. To render the transferred images we use pixel voting of the labelled pixels as presented in chapter 3.

The core of the algorithm (see Figure 4.2) relies on the metric used to compute the error. We evaluate the capability and robustness of three metrics to correctly associate

```
FUNCTION TransferByAnalogyMaterials(list_of_exemplars, photo-texture, &target_albedo, &target_shading)
{
    FOR (every_exemplar E) DO
    {
        coordinateMaps[E] <- TransferByAnalogy(E,photo-texture); // Trasnfer by Analogy
            error[E] <- computeError(coordinateMaps[E],E,photo-texture);
    }
        FOR (every pixel q) DO
        {
            finalCoordinateMap(q,E) <- minimumError(error[E],q); //finalCoordinateMap(_,E) defines the association pixel-exemplar
        }

    reconstructFromCoordinates(finalCoordinateMap, list_of_exemplars, &target_albedo, &target_shading);
}
```

Figure 4.2: *Transfer by Analogy Algorithm for several materials.*

a given pixel with the appropriate exemplar. We carry out several experiments with synthetic data and real data.

4.3.2 Evaluation of the Method

The TbA algorithm has several parameters: patch size, number of iterations and the metric used. Patch size influences the spatial coherence of the result and execution time. The number of iterations affects the fine grain of the matching. First, we wanted to evaluate the performance in relation to the metric used. We ran the experiment with patch sizes from 3 to 19 and found that the quality of the results is coherent for the three metrics across the patch size. We discuss the influence of the patch size in more detail in section 4.3.4. Similarly with the number of iterations. Therefore, for metric comparisons we fixed the patch size to 7×7 pixels and the number of iterations to 10. The three metrics evaluated are: Sum of squared differences (SSD) shown in equation 4.4a, Zero-mean Normalised Cross Correlation (ZNCC) in equation 4.4b, and a new Log Based Metric (LBM) detailed in equation 4.4c.

$$SSD = \sum_{x,y} (x - y)^2 \tag{4.4a}$$

$$ZNCC = \sum_{x,y} \frac{((x - \bar{x}) - (y - \bar{y}))^2}{\sigma_X \sigma_Y} \tag{4.4b}$$

$$LBM = \sum_{x,y} \log(1 + abs(x - y))^2; \tag{4.4c}$$

Where x, y are each of the pixels in the patches X, Y respectively. SSD is a commonly used metric used in texture synthesis. SSD has limitations in capturing the

perceptual difference between patches [HJO⁺01], but is still the most commonly used metric for reasons of simplicity and lack of more adequate ones. It is very sensitive to noise and also to large differences in a small number of pixels, because of its square factor. ZNCC is a widely used similarity metric for comparing small patches in computer vision. ZNCC is invariant to intensity and contrast, and has low sensitivity to changes in orientation. In our experiments, ZNCC showed a poor performance, proving not to be an adequate metric for our purpose. Due to its invariance to intensity and contrast, it loses important texture information and is unable to discriminate between exemplars. The objective of LBM is to encourage good matches, as in SSD, but not to penalise excessively large localised differences, favouring partial matching and robustness against noise.

(a) SSD
(b) LBM
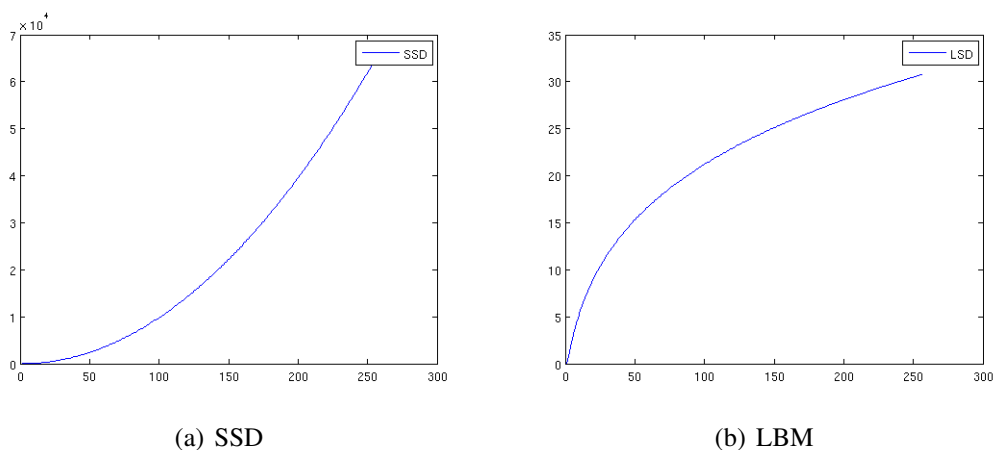
Figure 4.3: *Graphical response of the SSD and LBM. The x axis represents the difference in value between two pixels. The y axis shows the corresponding value returned by the metric.*

Figure 4.3 shows graphically the response of both SSD and LBM to the difference in a pixel. Analysing the graphs we find two main differences, the evolution of the graph with the increment of the value in the X axis – difference in value between two pixels – and the value returned by the metric in the Y axis. The value in the Y axis determines how important is the penalisation of a single pixel with respect to the total. In LBM, having a pixel that is totally different (255 in X), adds a value of 33 to the metric; which is equivalent to having two pixels with difference of 50, or six good matches where the difference is 10. SSD has a more radical behaviour. One pixel with difference 255 in X gives a value of 65025 in Y, which is higher than 25 pixels with a difference of 50. That means that a patch with perfect match for all the pixels except

two very different pixels, will be considered a worse match than one with all the pixels with a difference of 50. That makes SSD very sensitive to noise and partial matching, and LBM more robust to these two phenomena.

4.3.3 Synthetic Data Experiment

We designed an experiment to test the concept of our extended TbA algorithm for texture association. We created a synthetic image (Figure 4.5(a)) by combining different textures that had been created using exemplar based texture synthesis techniques [EF01, Ash01, Ash03]. We used the original exemplars(Figure 4.4), ran the segmenting algorithm, and created a false colour image that represents the exemplar that has been associated with each pixel.
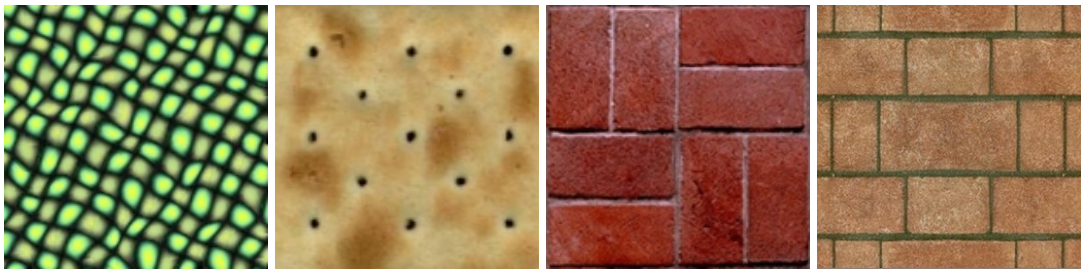
Figure 4.4: *Exemplars used for segmentation*

We fixed the patch size and the number of iterations so we could evaluate the different metrics only. In our first experiment we simply took an example with four distinct textures shown in Figure 4.5(a). In this simple example, SSD assigned 97% of the pixels correctly. Using LBM, our algorithm increased the correct association to the 99.8% of the pixels. ZNCC behaved poorly with close to 15% of wrongly labeled pixels. Since the textures are synthesised using the exemplars, we expected an almost perfect result for this experiment, only finding problems in the boundaries between textures. Surprisingly, SSD does not provide a close enough to perfect result. Analysing the result, we find that SSD is not able to associate properly patches that show small variations from the original exemplar. LBM is able to handle these small variations, performing more robustly. ZNCC is not able to discriminate the correct material.

In a second experiment, we ran the algorithm using eight exemplars, four of which did not correspond to any texture in the image. We aimed to evaluate the robustness of the associating algorithm when adding more exemplars. In the case of SSD, the error increased from 3% to 4.3%, so the accuracy was over 95% of the pixels. The percentage of correctly associated pixels with LBM remained above the 99%.

(a) Synthetic example (b) Ground-truth segmentation

(c) Resulting segmentation (SSD) (d) Resulting segmentation(ZNCC) (e) Resulting segmentation(LBM)
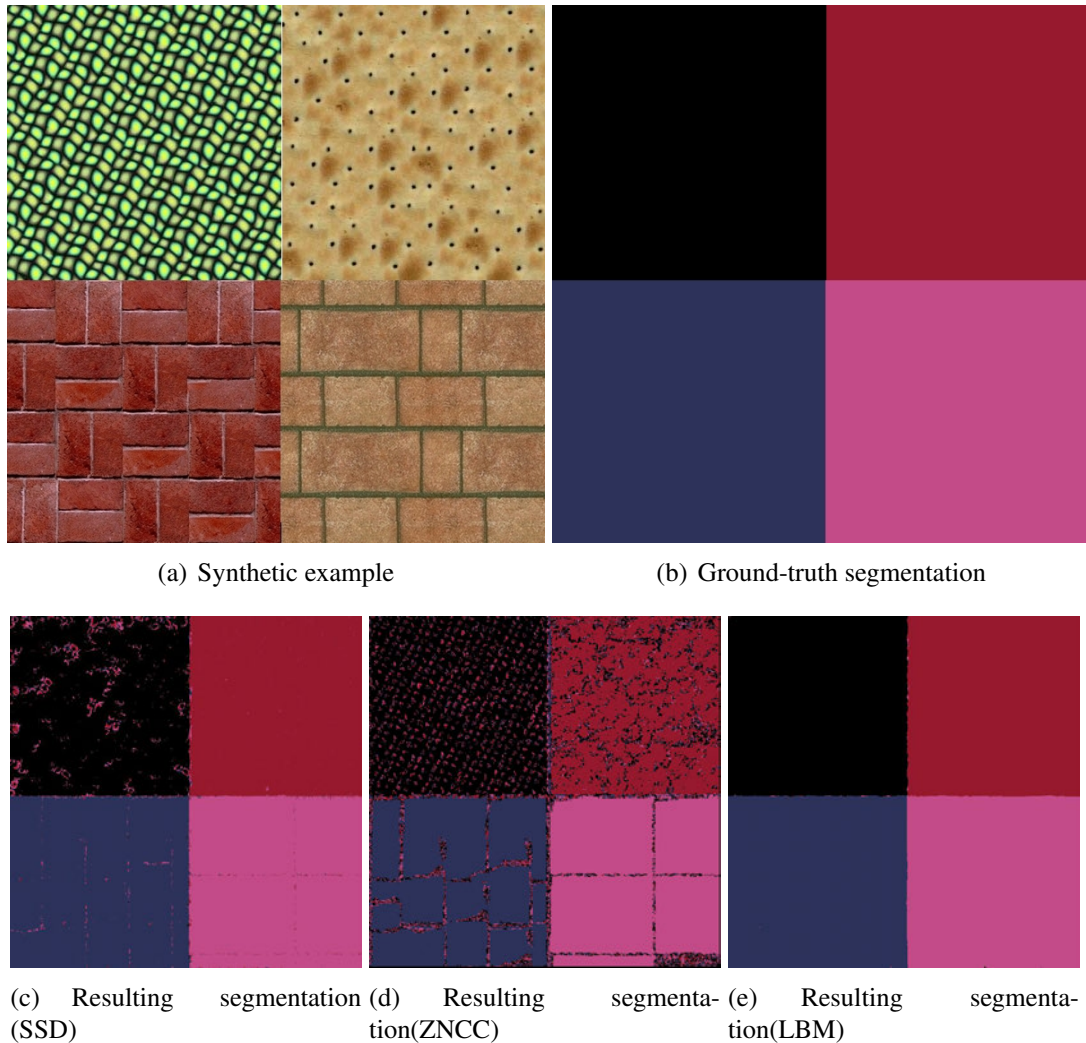
Figure 4.5: *The assignment is correct in the 97% of the pixels for SSD and 99.8% for LBM. Only the 85% of the pixels are correctly labeled with ZNCC*

The algorithm showed promising results in these tests, being able to associate a large percentage of the pixels correctly. We found that LBM performs better than the other two metrics tested. Finally, increasing the number of exemplars had a small influence on the result in this experiment .

4.3.4 Real Data Experiment

We have tested the idea of patch-based segmentation on different textures in the synthetic experiment. In this section, we tested the robustness of this approach on a real image of the façade of a building. This façade case is challenging since textures have

both great local variations and similarities between them.

To make evaluation and execution times tractable, we crop an interesting area in the façade in Figure 4.6, to obtain a section that contains several materials and refer to this as the *test image* (Figure 4.7(a)). Then we select small areas in the full façade, containing one type of texture or interesting features like windows or a water drainpipe, to be our 14 exemplars (see Figure 4.6). Some of these exemplar textures are contained in the *test image* and some are not.
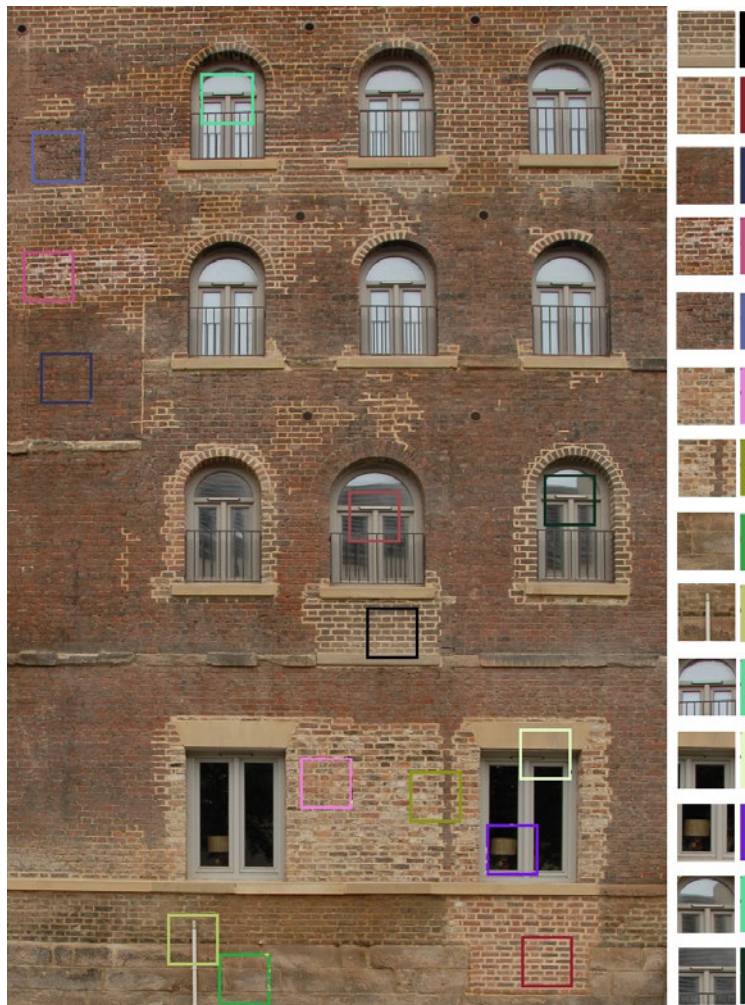
Figure 4.6: *Real Data Experiment: Reference Façade and Exemplars extracted from it.*

We first ran our algorithm using just the three exemplars in Figure 4.7(c). Exemplar 1 covers part of the test image. The texture in Exemplar 2 is not present in the test image, but we can see it shares some similarities with Exemplar 1 such as the colour between the bricks. Exemplar 3 is similar to the red brick in the test image. We show in

(a) Test image

(b) Manual Segmentation for reference

(c) Exemplars used for the first experiment
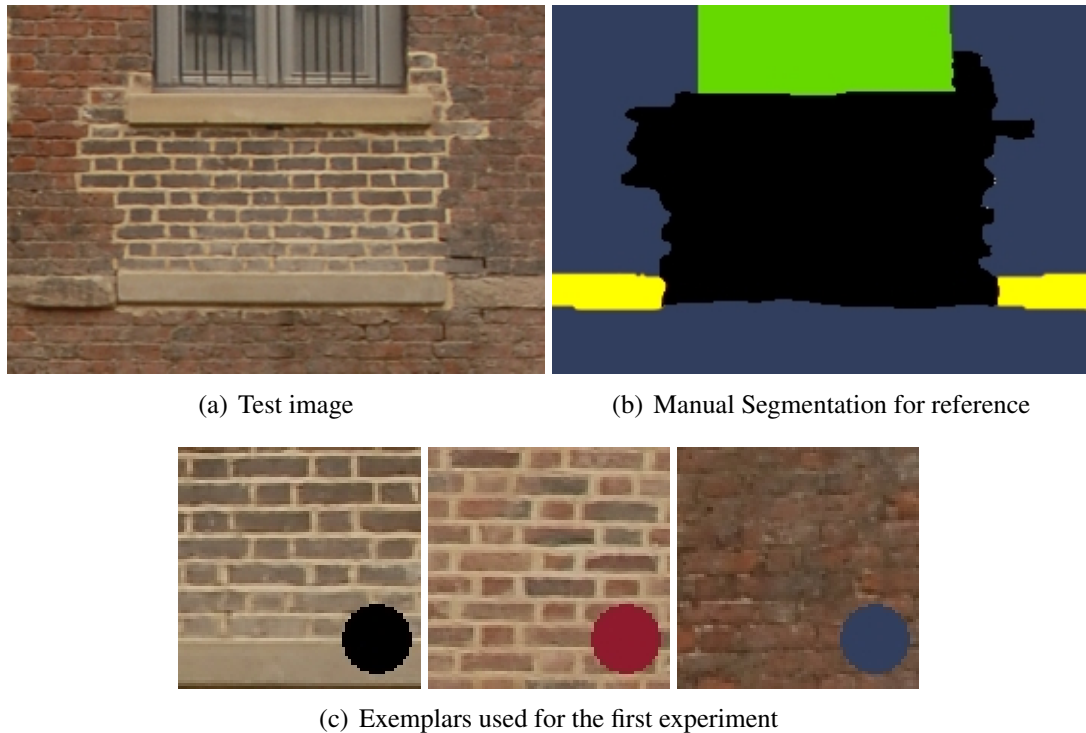
Figure 4.7: *Real Data Experiment: Reference Texture*

Figure 4.7(b) a manual labelling for reference. The reference labelling in Figure 4.7(b) gives an idea of how the association should look, although this is not necessarily "the perfect segmentation" for this data. Note that *YELLOW* and *GREEN* segments in Figure 4.7(b) are not associated to any material, since they are not contained in any of the exemplars. Exemplar 2 (middle in Figure 4.7(c)) is not associated to any segment in our manual segmentation. Exemplars 1 and 3 (left and right) are assigned a colour marked with a small circle in the bottom right corner. Unfortunately we cannot create a ground-truth labelling for the test image because the best association is the one that best transfers the underlying material characteristics. Therefore, we perform a visual evaluation of the results where we expect a coherent association at a global level and correctly matched local features.

The aim of this experiment is to test the two parameters, patch size and metric, with real data. We run the experiment with patch sizes from 5 to 19 (only odd numbers) and with the 3 suggested metrics: SSD, ZNCC, and our LBM.

Patch Size: Robustness, Behaviour and Relationship with Texture Features

Taking the test image and the three exemplars, we ran the the extended TbA algorithm and assigned a colour to the pixels according to the exemplar that presents the lowest value for the metric. The patch sizes are: 5,7,9,11,13,15,17, and 19 for all three metrics: SSD, ZNCC, and LBM. The patch size must be odd to create a symmetric square around the pixel. Figure 4.8 shows the results for 5, 9, 15 and 19, which provide a good illustration of the behaviour of the technique. These images follow the colour code of Figure 4.7, being areas in black, red, and blue labelled to exemplars 1, 2, and 3, respectively. The results should look similar to the reference segmentation except for the yellow and green areas.

(a) SSD

(b) ZNCC

(c) LBM
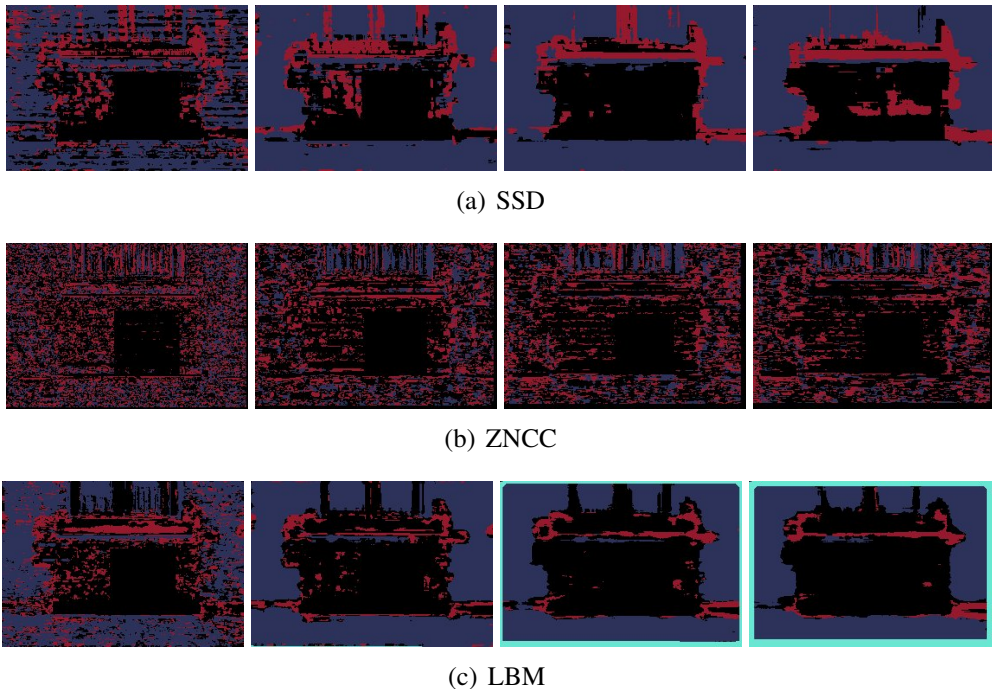
Figure 4.8: *Real Data Experiment: Pixel size and metric performance. 5-9-15-19*

The first observation on the results is that ZNCC does not produce spatially coherent results, even with big patch sizes, the matching produces noisy associations except for the squared area shared between exemplar and test image. This area is clearly distinguishable in the first left image of Figure 4.8(b).

The results from SSD vary with the patch size, gaining more coherent associations with bigger patch sizes. However, when the size is 17 or bigger, the metric starts getting correspondences with the wrong exemplar; even within the shared square area. This happens because, at the lower resolutions, the patches have pixels outside the

shared area and SSD finds wrong best matches, because of its sensitivity to noise. Then, because the exemplar is small, the algorithm is not able to correct for these bad matches. LBM is more stable with different patch sizes and gets a good solution more quickly with smaller patches.

From this experiment, we conclude that the patch needs to be big enough so it keeps the structure of the texture. If the patch is smaller than the texture features, the resulting associations are spatially incoherent. If the patch size is too big, then the metric has problems to find an appropriate feature, since it becomes difficult to find a good match. The patch size has a dramatic effect on the execution time, since the complexity of the algorithm is $N \times PatchSize^2$. Therefore, to be able to get a good association with smaller patch sizes is a useful property. We found that patch sizes of 7 and 9 behave well, providing a good performance in execution time and feature matching. LBM provided the most coherent results of the three metrics tested as it did in the synthetic case.

Number of Exemplars:Robustness

The second experiment aimed to asses the robustness of the algorithm when several exemplars are used. We fixed the patch-size to 9 and use the LBM, since this proved to work best. We increased the number of exemplars, one by one, up to the fourteen exemplars available, associating a new colour to every exemplar. The complete colour code can be found in Figure 4.6

(a) 3 4 5 6 7 8 exemplars

(b) 9 10 11 12 13 14
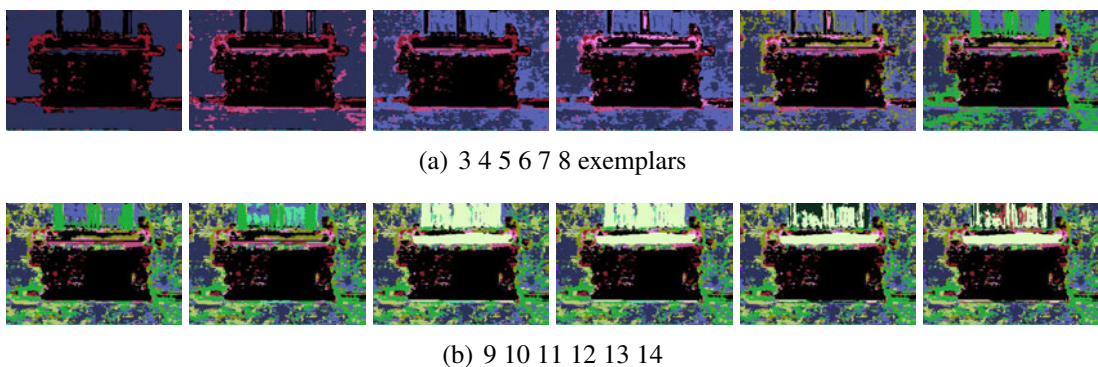
Figure 4.9: *Real Data Experiment: Robustness to the number of exemplars*

We observe in Figure 4.9 that the texture corresponding to the first exemplar (black) remains stable in the association. Also, when we introduce exemplar 11, cornice and window are well associated to it (yellowish green). When the material is not regular,

the method has problems to find spatially coherent associations, since the texture regions have more local variation than the exemplars. The interesting aspect of these results is that, when the exemplar is a good match, the labelling remains stable independently of the number of exemplars.

We finally ran the association algorithm with 3, 6 and 14 exemplars for the whole façade in order to evaluate the behaviour in a more complex example. We can see the results in Figure 4.10.
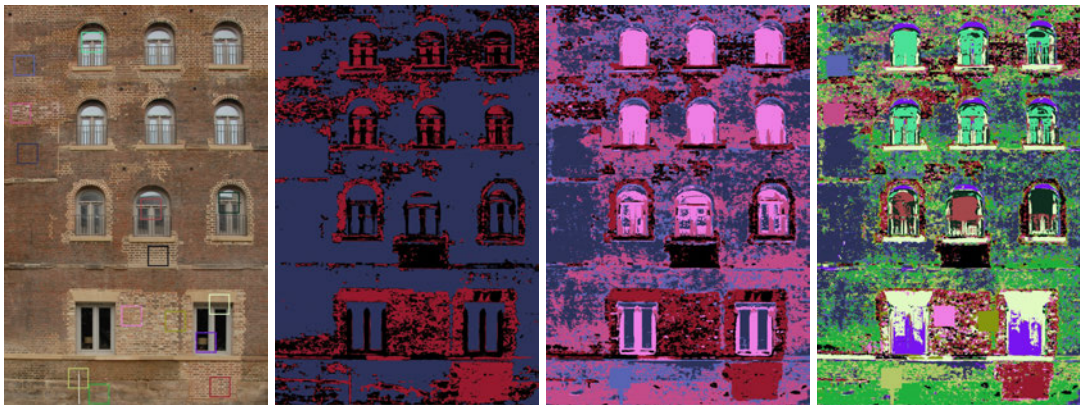
Figure 4.10: *Real Data Experiment: Robustness to the number of exemplars for the full façade. From left to right: Façade, association with 3, 6, and 14 exemplars*

The results for the full façade present similar behaviour to the partial image ones. With a small number of exemplars, the labelling could be useful for segmentation purposes adding spatial constraints similar to the ones presented in section 4.2. If the exemplars share features, the algorithm creates associations that mixed both exemplars. In the presence of many exemplars, the algorithm creates very fine grain associations. This could be problematic when using them for material transfer, producing a loss of structural information. Even in the presence of many exemplars, good material associations are consistent, for instance, all the cornices under the windows are associated to the same material.

4.3.5 Conclusion of the Experiments

The associations are correct locally, and our metric LBM produces more coherent associations than SSD or ZNCC. These associations are too fine grain to be used directly for segmentation, but we believe that they suffice for material transfer. Including many materials, especially when they are similar, reduces the spatial coherency of the associations and this can affect the structural coherence of the meso-structure transfer.

This automatic association technique is promising, but has some intrinsic limitations that are discussed in the next section.

4.3.6 Limitations and Future Work

Our association technique has two main limitations inherited from patch matching: scale and orientation. Beside these, our matching is performed under the condition of exemplars and façade captured under similar lighting conditions. Scale needs to be matched between exemplars and façade, and features only can be matched in the orientation in which they were captured, i.e. the orientation in the exemplar has to be the same as the orientation in the façade's texture. We capture exemplars and façades under diffuse lighting conditions, but the intensity of the lighting needs to be matched in order to be comparable. We discuss how we solve these limitations for our application and suggest future work for more automatic solutions.

The Intrinsic Limitation of Scale

We match scale manually between exemplars and the façade. This is a simple task in most cases, solved by measuring the number of pixels covered by a repeated pattern (such as a brick or a stone tile) in both the exemplar and the photo-texture, and finding the ratio between them. Then, we scale down the exemplar to match the scale of the façade.

We investigated briefly the possibility of matching scale automatically. Most work regarding scale in computer vision, tries to find scale-invariant features and descriptors for feature matching problems [LP92, DSH00, BL05]. This seems to work for image stitching or object matching, but require features to be present in both images. In our case, we want to match textures that are the same, but not necessarily contain the same image information, i.e. different samples of the same material.
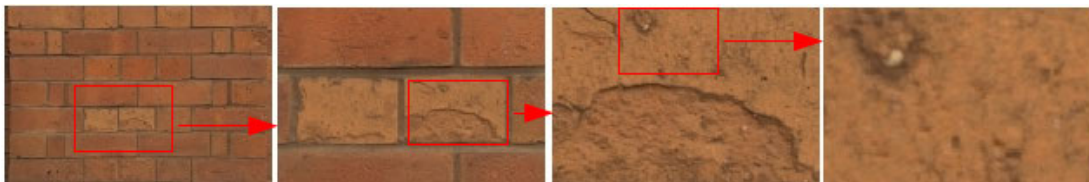
Figure 4.11: *Variation across scale. The same texture has different appearance at different scales.*

We experimented with the patch matching technique as a way of automatically

matching the scale. We select random samples in the texture and find the closest match in the exemplar at different scales. Computing the complete cost for all the patches, we have a measure of how well these two textures match for a given scale. If patches would represent appropriately the texture, then the scale with lower cost would be more likely to be at the same scale. We tested this hypothesis, but discovered that this method only works when the features sampled in one image are contained in the other. Unfortunately, when this is not the case, due to the variation across scale of textures illustrated in Figure 4.11, our technique was able to find good matches at different scales.

Recent work in this topic [HSNC08], seems to solve a similar problem and apply it to segmentation, by defining a scale descriptor for texture. Future work in this interesting topic could be applied in our pipeline.

The Intrinsic Limitation of Feature Orientation

Patch matching works on a unique orientation. This assumption normally holds for buildings, since bricks, stones, and other construction materials, are normally set in constant orientation. Normally we can capture the textures in this orientation, but sometimes, the orientation changes to make arches and other architectural features. Matching orientation locally implies the possibility for orientation changes in the patch, which would increase the number of comparisons per patch substantially.

We solve the problem partially by using an orientation field over the texture that deforms the neighbourhood and therefore the patch, adapting the orientation in the texture with respect to the exemplar. We use this to deal with deformations in the texture due to geometry, so we can apply our algorithms over complex geometries rather than only flat facades. This could be used to account for changes in orientation as well, but the challenge of computing an orientation field automatically remains to be explored. We discuss orientation issues in more detail in chapter 5.

Matching Ambient Appearance under Different Lighting Conditions

In the experiments presented in previous sections, exemplars were extracted from the façade photo-texture, so images are comparable. When exemplars are captured separately, images need to be normalised. To solve that, we crop a section of the facade similar to the exemplar and histogram match the ambient image of the exemplar to this selection. This compensates for any disparities in intensity, or colour lighting.

4.4 Multiple Material Transfer over change of scale

In this section we review the whole transfer process using a complete façade of a representative building in Manchester as an example. We capture four exemplars and perform the three combinations of the techniques presented here: Segmentation and Histogram Matching, Segmentation and TbA, and Automatic association and TbA.

4.4.1 Data capture and processing

We capture the two kinds of data required in our system: exemplars and a façade photo-texture. We identify four materials in the façade: red brick, sandstone, stone 1, and stone 2 captured using the depth hallucination method [GWJ+08] and shown in Figure 4.12. That provides four complete sets of high resolution images (2616×3900 pixels) containing an ambient capture, albedo map, shading map.
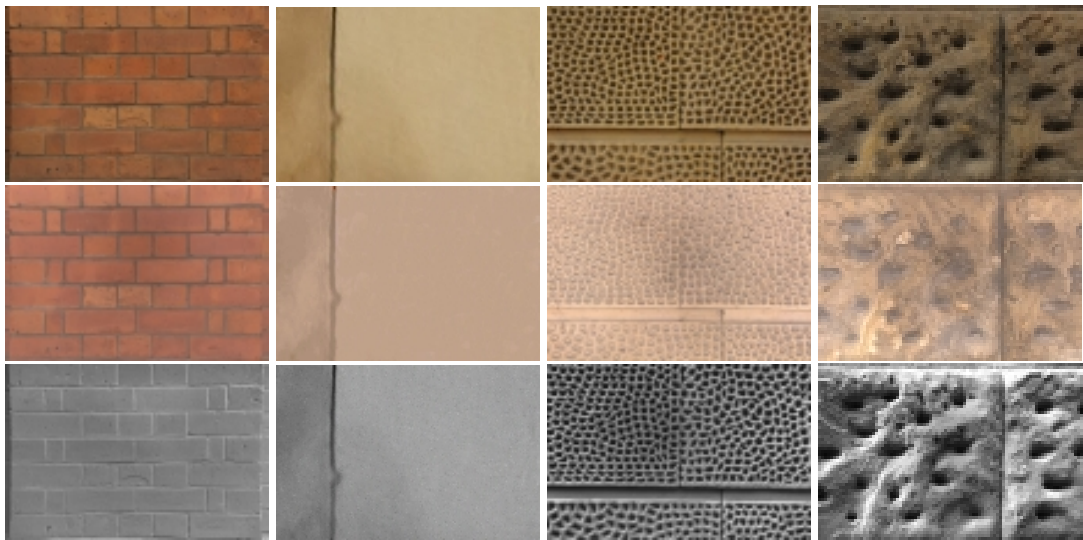
Figure 4.12: *(From left to right) Exemplars Captured Using SDH: Red Brick, Sandstone, Stone 1, and Stone 2. First row: Ambient capture. Second row: Albedo map, Third row: Shading map.*

The second data source is a photo-texture. By combining together several high resolution images we manage to get a 9126×9856 pixels photo texture (Figure 4.13) [1].

The elements in the picture that we are not interested in, or we do not have exemplars for, such as cars, windows, background, and doors, are segmented out.

---

[1] Image created by Timo Kunkel. http://www.cs.bris.ac.uk/ kunkel/

Figure 4.13: *Photo-texture reconstructed from several images of a building in Manchester*

4.4.2 Material Transfer by Histogram Matching

We created a file that defines the association of every segment with its corresponding exemplar. We then histogram matched every segment separately, and merged all the histogram matched segments in a single image. This provided a full resolution albedo map (Figure 4.14) and shading map. We computed the depth map using the depth-hallucination technique and a normal map from the depth map.

Looking closer at the albedo map and the normals computed from the transferred

Figure 4.14: *Resulting Albedo map Using Segmentation and Histogram Matching*

surface detail (see Figure 4.15), we can see the quality of the results. The appear-
ance of the four materials is preserved and also the surface detail. This is difficult to
evaluate, since we do not have ground truth, but visually, we can see the plausibility
of the results compared with the captured albedo maps for the exemplars and with
the expected geometry from the original photo-texture. An interesting characteristic
of this technique is that it keeps the geometry from the original image. For example,
the ornaments in the gate are well captured as well as the broken areas in the *Stone 1*

material.

(a) Close up detail of the Albedo Map

(b) Exemplar Albedo Maps for comparison

(c) Close up detail of the Normal Map

Figure 4.15: *Close up views of Transferred maps using Histogram Matching. Albedo is consistent with the albedo maps captured for the exemplars. Geometric detail preserves the local detail consistently with the appearance of the façade.*

4.4.3 Material Transfer by Analogy

We ran the experiment using the same segments, and the same exemplars and associations. We standardised the exemplars as presented earlier, by cropping a piece of the photo-texture that contains the same material in the exemplar, and then histogram match the exemplar ambient map to this image. The result is an exemplar ambient

map that matches the lighting conditions of the material in the façade. If the lighting conditions for a material change over the photo-texture, we need to perform this standardisation for every segment that is captured under different lighting conditions. Albedo and shading maps remain unmodified.

We apply our TbA technique to the segments. We standardised the red brick material for every floor in the façade, since in the photo-texture the higher floors present a brighter appearance than the lower floors, due to not completely diffuse lighting conditions when capturing the photo-texture for the façade. Under completely diffuse lighting condition, like an overcast day, this process is not necessary.

(a) Close up detail of the Albedo Map

(b) Close up detail of the Normal Map

Figure 4.16: *Close up views of Transferred maps using Transfer by Analogy. Albedo is consistent with the albedo maps captured for the exemplars. Geometric detail preserves the local detail consistently with the appearance of the façade.*

The results are similar in quality to the ones produced by Histogram Matching. The final maps lose some level of detail compared to the Histogram Matching ones, but the technique is still able to produce good results. In areas where the features are not present in the exemplar, like the ornaments on the entrance, the process still results in acceptable albedo and normal maps. The application of this transfer technique does

not provide any advantage in this case, since we still carried out the segmentation process, and an extra exemplar standardisation process. Also, the materials are globally regular, so the advantage of local matching that TbA provides, does not provide an improvement in the result compared to Histogram Matching. Besides that, TbA is able to transfer albedo and shading appropriately.

4.4.4 Automatic Association and Transfer by Analogy

The final experiment is to run the association process that we described in section 4.3. The resulting association is shown in Figure 4.17. We can see that compared with the manual and semi-automatic segmentation, the association is at a similar level of accuracy, although not perfect. Brick and sandstone in the window frameworks are well associated. Ornaments in the door and windows are not associated to sandstone since the exemplar does not contain features of this kind. There are some areas in shadow, with complex features or albedo changes, where the matching is not completely correct, for example under the cornices or in the eaves. Taking into account that this process is completely automatic, the results are very encouraging for further research, but the level of accuracy required by a segmentation makes an interactive approach more suitable.

(a) Manual   (b) Our interactive approach   (c) TbA Segmentation
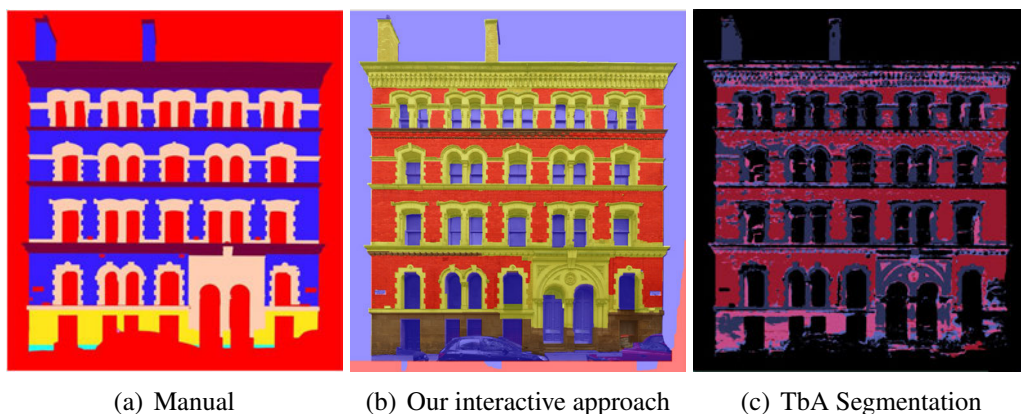
Figure 4.17: *Comparison between the three segmentation methods*

The albedo and shading images are reconstructed by weighting the albedo and shading maps from the exemplars according to the value of the metric. Figure 4.18 shows the detail of a section in the façade containing three of the four materials: *Red Brick, Sandstone,*and *Stone 1*. This section is automatically associated and transferred using our extension of TbA.
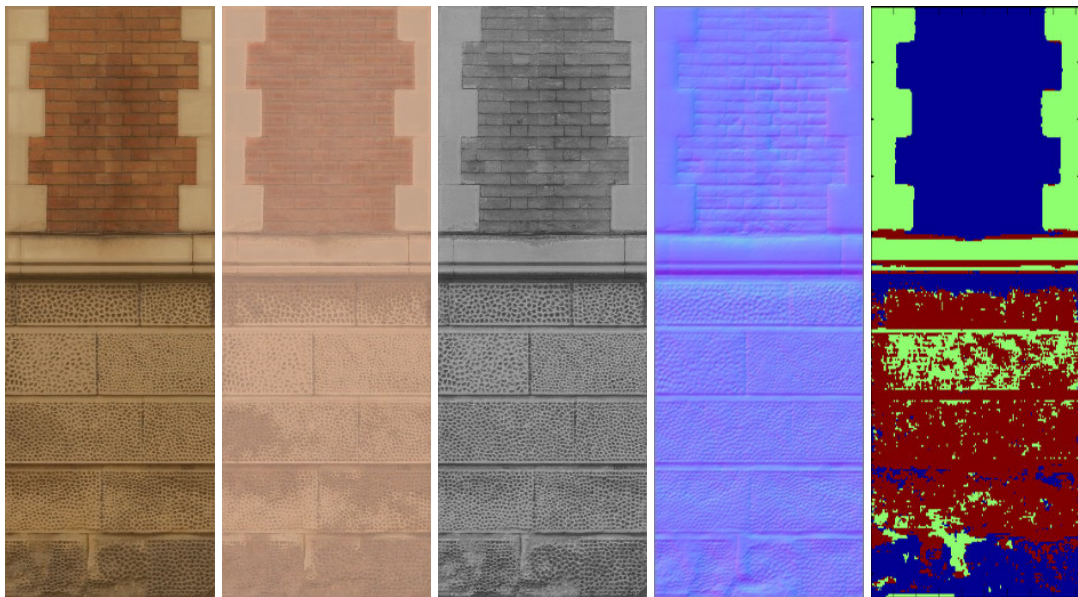
Figure 4.18: *Automatic association and Transfer by Analogy. (From left to right) Original section of the photo-texture. Transferred Albedo, Transferred Shading Map, Normal Map, and Automatic Labelling*

The final result is plausible in these examples. Some resolution is lost as result of the Transfer by Analogy process. The association image (last in the right), shows that the red brick (in blue) and the sand stone (in green), are correctly labelled. However, some areas with the stone 1 material are wrongly labelled. This is due to the some very light bricks in the red brick exemplar, and to some non-textured areas in the case of sand stone. This produces certain colour shift in certain areas of the albedo. The colour in the albedo is also partially washed out, for example in the brick area, due to the image creation process. This process weights the different materials according to the value given by the metric. In this case, this gives some yellow appearance to the albedo of the *brick* given by the *sandstone* and *stone 1* materials.

The shading map on the other hand, is well transferred. The texture structure of the three materials is well preserved and the final normal map is completely plausible.

We conclude from this experiment that the complete automatic process has limitations to find the correct label configuration. The weighted reconstruction process overcomes some of these limitations, particularly well in the case of the shading, but can produce artifacts in the albedo map.

Results are promising, and encourage further work on the automatic multi-material transfer process.

4.5   Conclusions and Future Work

We have presented solutions for the problem of transferring materials to large surfaces containing several materials for both transfer techniques. Histogram Matching can be applied to complete façades by segmenting, with manual or interactive techniques, the different materials, and apply the transfer method separately to the different segments. We implemented an interactive segmentation technique that produces suitable results, reducing the time necessary to perform the segmentation. Transfer by Analogy, can be applied similarly to the same workflow, producing similar results. This requires matching scale and lighting between exemplars and the photography of the façade. The complete workflow is operational and produced good albedo and meso-structure from a set of exemplars and a fronto-parallel view of a building façade.

Aiming for a completely automatic process, we investigated the extension of Transfer by Analogy to several materials. We evaluated several matching metrics, the novel Logarithm Based Metric being the one providing the best results. We analysed the effects of changing the patch size and the number of the exemplars on the behaviour of the labelling process. A size of 7 or 9 gives an appropriate trade-off between performance and execution time. The number of exemplars affects the result and can make the technique unstable, however, good matches normally remain stable.

The application of this technique to automatically transfer material properties to complete façades has produced promising results, but requires further investigation. Here, we investigated the use of data driven descriptors, which have been applied successfully to texture synthesis, and we conclude that they have limitations in robustness and accuracy. Our technique can take advantage of new advances in texture analysis, and deeper study of texture models and feature descriptors is suggested as future work.

Matching scale of textured surfaces is an interesting open research problem, and advances in this area may produce a significant improvements to our results.

Full Model Reconstruction

THIS *chapter deals with the three remaining problems in our reconstruction system. First, to reconstruct integrated texture maps for large structures. Second, to perform our texture characteristics transfer over a non-planar gross-scale geometry. This involves certain deformations of the target space that need to be taken into account. Third, to combine the transferred geometric detail with the gross-scale geometry previously reconstructed.*

5.1 Multi-view Texture Reconstruction

Texture reconstruction refers to the process of creating an optimal texture map from the set of perspective images used for the reconstruction of the gross-scale model. The definition of an optimal texture map depends on the application domain, but there are certain characteristics that are generally desirable: a compact representation, high image quality, and minimisation of artifacts. Besides these three, it is also important that the texture map has low distortion and has a continuous representation in order to apply our meso-scale and reflectance transfer techniques.

5.1.1 Compact Representation, Distortion, and Continuity: Parameterisation

A texture map is a 2D space where each vertex in a mesh has a position defined by its texture coordinates, normally denoted (u, v). This one-to-one mapping of the surface to the plane is termed *the parameterisation* and determines the distortion, continuity, and compactness of the texture. An example of planar parameterisation is an earth globe map, where the globe is mapped to a 2D plane. An ideal parameterisation of a triangular mesh is one that:

(a) Triangular mesh   (b) Texture atlas 1   (c) Texture atlas 2
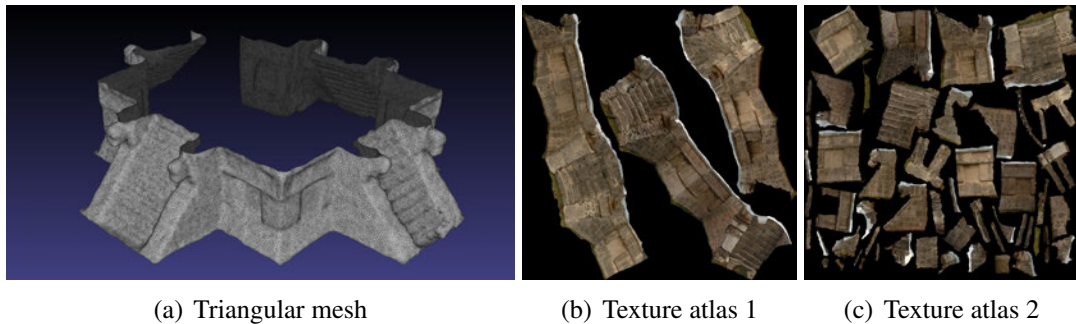
Figure 5.1: *Allowing larger deformations produces larger charts as in (b). Texture atlas 2 (c) shows a closer to isometric parameterisation at the cost of a more complex texture atlas.*

- Does not distort the triangles, so their angles are preserved (*conformal*) as well as their areas (*equiareal*).

- Maintains the continuity defined by the mesh, so that the neighbourhood topology of vertices is the same.

- Only the information required for texturing is stored.

A mapping that is conformal and equiareal is denoted *isometric* and only exists in very special cases [FH05]. Most algorithms for surface parameterisation attempt to find a mapping that is conformal, or equiareal, or minimises some combination of the two. A complete review of parameterisation algorithms can be found in [SPR06, HPS08].

The traditional approach to finding a low-distortion mapping for complex meshes is to cut the mesh into *charts*, which are parameterised individually, and then to pack them into a *texture atlas*, as shown in Figure 5.1. This allows creation of isometric, or close to isometric, mappings at the cost of cutting the mesh. Our texture reconstruction technique is independent of the parameterisation, so can take advantage of new advances in the area, but does require minimal distortions and, therefore, it needs to handle multiple charts.

Division of the mesh into charts introduces three further issues: (1) the mesh cutting problem; (2) the discontinuities between charts for geometric processing, and (3) the discontinuities in the parametric domain for texture processing.

The first two problems have attracted attention in the research community and several solutions have been proposed [EDD+95, GKSS02, LSS+98, PSS01] where the

challenge is to obtain mappings that are smooth across the patch boundaries. The cuts and parameterisations used in this thesis have been produced using the software GRAPHITE [GRA09]. This software includes some state-of-the-art parameterisation and cutting algorithms. It automatically creates a texture atlas from a given geometry allowing the customisation of several parameters such as maximum distortion, maximum scale, and minimum size of a patch.

The third problem is the discontinuity of the texture map which needs to be tackled when reconstructing the texture. Discontinuities can incur visual artifacts like *seams* if an edge shared by two charts is textured with different images, and also any texture processing that uses neighbouring information will require the original continuity to be maintained. Our proposed solution is to use an indirection map. This map is an interface between the user and the texture map masking the discontinuities introduced by the cutting process.

A similar approach was proposed by Lefebvre and Hoppe [LH06], and Gonzalez el al. [GP09], who used a level of indirection to avoid seams when accessing on-the-fly texture maps divided into charts. We, instead, use the indirection map during the texture reconstruction and during the texture processing to maintain continuity of the neighbourhood established by the mesh.

Indirection Map

The parameterisation defines the mapping of the 3D geometry to a planar space. This planar space is discretised to the resolution of the texture map. When allowing cuts, the empty space left between charts results in empty spaces in the texture map breaking the continuity, i.e. a neighbouring pixel in the texture map does not necessarily correspond to the neighbouring pixel mapped to the geometry.

To maintain continuity and produce seamless textures over a discontiguous atlas we create an indirection map. This has the same dimensions as the texture map, and links every position in the map with the appropriate pixel in the texture map (texel). The user accesses the map in the same way as he would access a continuous image and the indirection map masks discontinuities, blank spaces in the texture map, and solves problems at the boundaries.

There exist two types of texels in the texture map: interior (green and blue in Figure 5.2 (middle)) and exterior (grey and yellow). Interior texels lie inside the chart (the centre of the pixel is inside). Exterior texels correspond to the empty space between

charts and are not used for texturing. A subset of exterior texels lie on a boundary (yellow texels). These texels are exterior but have a neighbouring texel which is interior. Since we allow for cuts in the geometry for parameterisation, we can distinguish between geometric boundary texels (pink in Figure 5.2 (right)) and chart boundary texels (red and orange). Geometric boundary texels are those where the boundary is the end of the geometric model. Chart boundary texels are those where the boundary is shared by two charts.
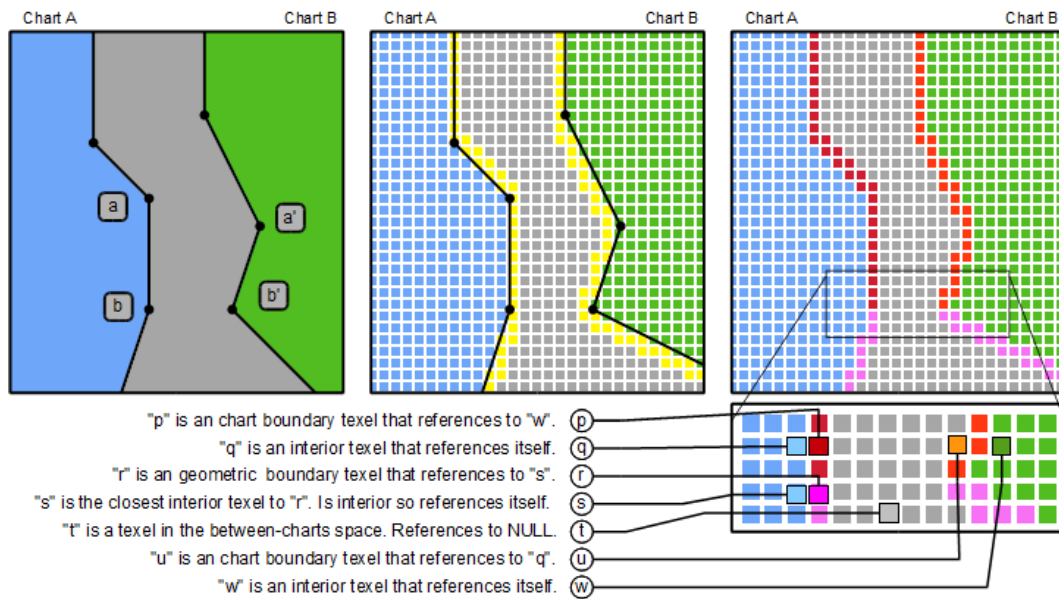
Figure 5.2: *Chart A and B shares an edge (a-b) in A and (a'-b') in B). The space between charts is coloured grey. The image in the middle shows the discretisation in texels, and the boundary texels marked in yellow. These texels are Geometric boundaries(red and orange) and chart boundaries (pink).*

We denote a texture map $T[]$, and the indirection $S[]$. For a position p, $S[p]$ returns the position of the appropriate texel in the texture map. There are four possible indirections:

1. If the texel p is valid and interior, the position referenced is the same. $T[S[p]] = T[p]$

2. If the texel t is exterior and not in the boundary, the indirection map acts as a mask. $T[S[t]] = NULL$

3. If the texel q is a chart boundary, the indirection map references w, which is an interior point in the neighbouring triangle in the appropriate chart. $T[S[q]] = T[w]$

4. If the texel r is a geometric boundary, the indirection map references to the closest interior and valid texel s. $T[S[r]] = T[s]$

During the construction of the indirection map, boundary texels are detected when discretising the mesh in the texture space, then assigned to the closest interior texel. In the case of a chart boundary texel, the boundary point has a corresponding texel in a different chart. We locate this texel by finding the shared edge and interpolating the position between the texels corresponding to the vertices of the edge. This position must be a boundary texel or a valid interior texel. In the case of a boundary texel, the closest interior valid point is assigned instead.

The resulting indirection map assigns every position in the map to a valid interior texel or to NULL for invalid empty texels. The indirection map also recovers the continuity and neighbourhood defined by the 3D geometric model within the 2D texture.

5.1.2 Texture Sampling Quality

The sampling quality of the reconstructed texture depends on the quality of the original images. Since a portion of the surface will be visible in several views, not all of them will provide the same quality, hence a quality metric needs to be defined to select the best view for every part of the texture. Gibson et al. simply used the distance to the camera [GHCH03]. Debevec et al. and, Lempitsky and Ivanov, used measures based only on the angle between the normal and the camera [DYB98, LI07]. Bernardini el al. included both factors [BMR01]. Several authors, used the projected area of the texel at every view, which takes into account projection, pose, and distance [POF98, Bau02, RAKRF08].

More formally Ismert el al. [IBG03] defined the texture quality metric of a view as a sampling problem.

> "The image formation process is represented as a mapping of the texture onto objects, then onto the image sensor. This mapping can be described by the image transform M_{Img}, which maps texture space to image space ($R^2_{(s,t)} \rightarrow R^2_{(u,v)}$). M_{Img} is composed of M_{Proj}, which projects the object into the image plane and is defined by the Projection Matrix P, and M_{Tex} which projects the surface onto the object that is defined by the parametrisation discussed in section 5.1.1. This transformation can be characterised by its Jacobian matrix $\mathbf{J}(\mathbf{M}_{\mathbf{Img}}^{-1})$:

$$J(M_{Img}^{-1}) = \begin{bmatrix} \frac{\partial s}{\partial u} & \frac{\partial s}{\partial t} \\ \frac{\partial t}{\partial u} & \frac{\partial t}{\partial v} \end{bmatrix} \quad (5.1)$$

Recall that $J(M_{Img}^{-1})$ maps (s,t) to (u,v). The values of the Jacobian matrix, as partial derivatives, indicate the change in sample distances in the directions indicated. Thus, the elements of this matrix describe the change in the sampling behaviour induced by the transformation. Note that because it is derived directly from M_{Img}^{-1}, $J(M_{Img}^{-1})$ accounts for all of the factors affecting the sampling rate, i.e. projective effects, camera pose or position, lens distortion, etc.

Evaluating the Jacobian matrix across the texture gives four measures per texel. Values in the Jacobian matrix > 1 indicate that the texture space in discretised more densely than the projected image space, while values < 1 indicates the opposite. To keep the detail contained in the original images, these values must be ≥ 1 at every texel."

By enforcing this condition we can determine the optimal resolution of the texture map.

The determinant of the Jacobian matrix $J(M_{Img}^{-1})$ is in fact the reprojected area of the texel in the image space as was used in [POF98, Bau02, RAKRF08]. This is a complete quality metric per texel and view, that accounts for all the transformations in the image process including camera angle to the view and distance. Our texture reconstruction process uses this metric to determine the best available view for every texel in the texture map.

5.1.3 Texture Artifacts

Apart from the sampling quality of the texture, there are other factors that determine the overall quality of the texture. We refer to these as artifacts that should be minimised. We identify four different types of artifacts:

- *Seams* appear when different images are used for texturing contiguous areas, due to differences in the lighting conditions, camera parameters, etc. We denote regions textured using the same image as *patches*.

- Occlusions due to the same object produce areas textured with either wrong, or no, information.

- Occlusion due to non-modelled objects which are present in the images, eg. a car or a person.

- *Ghosting* and *blurring* due to imprecisions in the model or blending of different views.

Reducing seams has received most attention in the literature on multi-view texturing. For example, a similar problem occurs in the process of image mosaics and panorama stitching, where images are taken under different conditions and need to be merged seamlessly. Solutions in the context of multi-view texturing usually blend the overlapping areas of the different views. Different weights are given to the different views according to a quality or confidence metric, as explained previously [DYB98, BMR01, POF98, Bau02]. The main problem with blending is that errors in alignment and big differences between neighbouring images (for instance when a person is present in one image but not in the other) produce blurring and double images or ghosting effects. To solve the blurring problem [Bau02, YLHS05] used a multi-band blending algorithm which merges separately high frequencies and low frequencies, preserving image detail and reducing the blurring. This algorithm was previously introduced by Burt and Adelson [BA83b], also referred as *Laplacian pyramid blending* or *multi-resolution spline blending*, applied to image mosaics.

An alternative technique for merging images seamlessly is to perform the combination in the *gradient domain*. This was first used in the 2D mosaicing problem by Agarwala et al. [ADA+04] and adapted to multi-view texturing in the 3D domain in [LI07]. Besides the effectiveness of combination techniques, seams need to be reduced to a minimum, which suggests minimising the number of patches.

Our new approach to texture reconstruction minimises both seams and number of patches using *graph-cuts* [BVZ01] based optimisation over a *Markov Random Field* [SZS+08b]. We define the texturing from multiple views as a mosaicing problem directly over the texture domain, including both quantitative and qualitative factors in the formulation. The remaining seams are removed using a gradient based method described in more detail in section 5.2.3.

Occlusions due to the same object need to be detected so the surface is not textured with a view where it is not visible. That involves a visibility test for every view and triangle in the model. Those areas that are not visible will need restoration, by using

texture synthesis or in-painting techniques. We in-paint small sections as described in section 5.2.3.

Occlusions due to objects that are not modelled are more difficult to handle, since we do not have evidence of the occlusion. Bornik et al. used a median filter that discards pixels that are more than 10% different to the average colour of the pixel in all the images to find non-modelled occlusions [BKBF02]. Similar procedures have been used to remove specular highlights which are also view-dependant [OSRW97]. This process will discard outliers as long as there exists sufficient evidence of the real colour of the surface, i.e. the colour of the pixel is similar in several images. Our implementation uses the same method to detect occlusions due to non-modelled objects.

In the next section we describe the texture reconstruction method developed and implemented during this thesis which is inspired by previous work but adapted to the requirements of our specific problem: reconstructing high resolution textures from automatically calibrated, wide-baseline, multi-view images and a gross-scale geometric model for texture transfer purposes.

5.2 Markov Random Field Texture Reconstruction

Optimal texture reconstruction from multi-view imagery requires a trade-off between the best sampling quality and the reduction of artefacts. Simply using the best view can incur artifacts like seams and occlusions due to non-modelled objects. On the other hand, minimising only the seams might produce badly sampled areas. We decided to formulate the problem within an optimisation framework where the different factors are included in the process.

We designed and implemented a novel solution that formulates the texture reconstruction problem in an optimisation framework, inspired by the work on image mosaics, over a Markov Random Field (MRF) defined on the texture space. MRFs have been successfully used in energy minimisation problems [SZS$^+$08b] in many image processing tasks in 2D like image segmentation [RKB04], selective compositing, 3D like mesh segmentation [LW08] and, more pertinently, image mosaics, panoramic stitching [ADA$^+$04], and multi-view texturing [LI07]. In fact [LI07] is very close to our formulation, but there are several important differences.

Our method performs the optimisation in texture space instead of geometric space, and so is independent of the mesh resolution. This way, we can control the resolution

of the optimisation by simply modifying the resolution of the MRF, which is easier than doing so in the mesh basis. Also, we can perform the optimisation on a texel basis, instead of a triangle basis in [LI07], which is important for occlusion detection. Furthermore, a 2D approach simplifies the possibility of including interactive techniques in the process similar to those in [ADA+04], in-painting, and other image processing algorithms. The method of Lempitsky et al. performs the optimisation directly in 3D avoiding parameterisation and therefore distortions, However, since we need to perform other processing in the texture domain, parameterisation is necessary, and optimising in 3D is of no advantage to us.

Texture reconstruction as a 2D optimisation problem shares its three main steps with the mosaic generation process. First to find a common parametric space for all views is necessary. In this case, the common space is the texture space. By re-sampling every view in the texture domain, we set all the images in the same space. Second, we need to choose the best view for every texel. This process is performed by minimising an energy function that encodes both reduction of seams and sampling qualities. The final step is to remove the remaining seams. As mentioned before, we use a gradient based method for this purpose. In the following sections, we present the details of our solution.

5.2.1 Finding a Common Parametric Space

The first step in the texture reconstruction process is to parameterise the geometric model in a planar surface with minimal distortion. Once we have the model in a flat chart mosaic, we choose a resolution for the texture and scale our parameterised model accordingly. Then we discretise the model by sampling in equal unitary intervals the geometry, creating a 2D regular representation of the geometry, where every position in the grid stores a three dimensional vector with the 3D geometric coordinates. Having the position of the geometry for every pixel, we sample every view for every point in the geometry image acquiring a texture map for every view, all of them in the same texture space. Figure 5.3 shows three texture maps corresponding to three different views.

Unifying all the views in the same parametric space allows us to choose the best view to texture every texel using our optimisation framework.
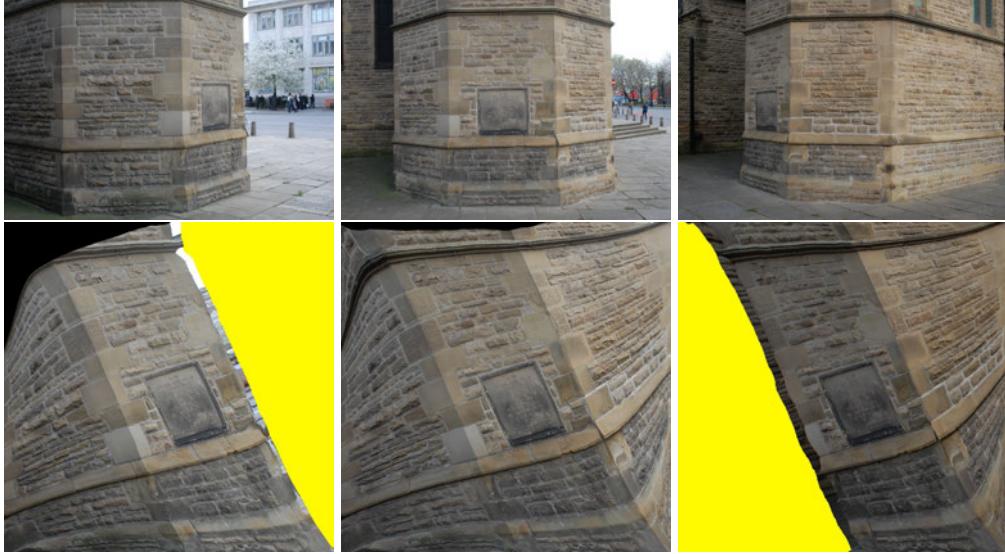
Figure 5.3: *(Top row) Three different views used for texture reconstruction. (Bottom row) Corresponding texture maps in a unified parametric space. Yellow means this area is not visible in this view.*

5.2.2 Camera Selection as an Optimisation Problem

A Markov Ramdom Field (MRF) is an undirected graph consisting of a set of nodes and a set of links between nodes, and contains a probability or cost associated with every node and link. In our case, each node is a texel, and the objective is to find the most appropriate camera to texture each node, minimising the seams, and maximising the sampling quality. These two factors to be optimised are encoded in an energy function established by the equation

$$E(C) = E_Q(C) + \lambda E_S(C) \qquad (5.2)$$

for a set of possible cameras $C = \{c_1, c_2, ..., c_n\}$. This equation contains two terms (a data cost term $E_Q(C)$ and a smoothness cost term $E_S(C)$) and a scalar factor λ. The term $E_Q(C)$ assigns a penalty for using a camera to texture a texel. This penalty is inversely proportional to the sampling quality for this camera, the better the quality the lower the cost. Visibility is also encoded in this term by setting the cost to the maximum value when the node is not visible in this camera. The second term $E_S(C)$, defines the cost of assigning different cameras to neighbouring nodes. This term encodes the seams. If both neighbouring texels are labelled with the same camera, the penalty is zero, since there is no seam. When the cameras are different, this penalty is proportional to the difference between the texels. The factor λ is a parameter

defined by the user to give more importance to the cost or the smoothness term. A high value of λ prioritises the smoothness term giving a solution with larger areas textured with the same camera.

The data cost term is defined as $E_Q(C) = \sum_t C_d(t, c_i))$ where C_d specifies the penalty of assigning the camera c_i to a texel t. This penalty is the inverse of the projected area of the texel t in the image i in pixels. Since the area of the texel is unity, the projected area in the image i is equivalent to the determinant of the Jacobian as explained in the section 5.1.2. Two visibility tests are also performed. A self occlusion test that works by re-projecting every texel in the original image and producing a depth map using the distance from the point in the geometry at this texel to the camera position, and a second visibility test for occlusions due to non-modelled objects. This uses a median filter, as in [BKBF02], that compares the value of this texel for a specific camera with the average value of this texel, computed with all the available cameras.

The second term is defined as $E_S(C) = \sum_{t,s} C_s(t, s, c_i, c_j)$ and encodes the seams due to change of camera. C_s specifies the penalty of assigning a camera to neighbouring texels t and s with the cameras c_i and c_j respectively. If $c_i = c_j$ then the penalty is $C_s() = 0$ since there are no seams. If $c_i \neq c_j$ then the penalty $C_s()$ is the Euclidean distance in RGB colour space between texel t in camera c_i and texel s in camera c_j. When accessing neighbouring texels we use the *indirection map* as described previously. This preserves the continuity of the texture through the geometric model avoiding the spaces between charts.

The parameter λ simply specifies the priority assigned to quality or to seam reduction and is specified by the user. Assigning $\lambda = 0$ will use the best sampling quality camera for every texel. On the other hand, high values of λ will give preference to minimising seams with respect to the texture quality, providing a solution with fewer cuts and therefore larger areas textured with the same camera.

While the MRF framework yields an optimisation problem than is NP hard, good approximation techniques are available [SZS+08b]. We use graph-cut optimisation [BVZ01]. We typically create a high-resolution texture mosaic of 7000×7000 pixels, although the process allows for arbitrary resolutions, limited only by the memory of the system. During the optimisation process we instantiate in memory the data cost maps for every view. Since the number of views can be large, we run the optimisation at a lower resolution, depending on the number of views, and then we scale up the association map and sample the texture map at full resolution.

5.2.3 Removing Remaining Seams: Poisson Reintegration

The last step in the texture reconstruction process is reducing any remaining seams. We use gradient based fusion similarly to [ADA+04]. The algorithm is based on solving the *Poisson equation* using *Neumann* boundary conditions i.e. dropping any equations that involve pixels outside the boundary of the image. This method can reconstruct the original images from their gradients up to a scale factor setting all the images to the same global brightness level, performing a global correction in the colour. By modifying the gradient field, this method can be used to perform several image editing processes, such as inserting objects in pictures seamlessly [PGB03].

Figure 5.4: *Remaining seams due to lighting variation shown in the top row, can be totally removed using poisson reintegration as can be observed in the bottom row. The middle and right columns correspond respectively to the areas inside the red and blue squares highlighted in the right column images.*

We use this technique to remove any remaining seams by reintegrating the image from its gradients. We create a gradient field that takes the gradients from the different views. Then, we apply the Poisson Reintegration to create a seamless texture map. Figure 5.4 shows a texture map reconstructed with our technique. The top row shows some remaining seams which are removed using the gradient fusion shown in the bottom row. In general, the remaining seams are small, like this example, because the lighting conditions do not vary a lot in the short time required to capture the wide-baseline sequence. However, the technique is also able to deal with more severe seams as demonstrated by Perez et al. [PGB03].

Figure 5.5: *Our implementation of the Poisson reintegration algorithm can also in-paint small areas and remove shadows. In the example of the left, a reconstructed texture, using our system, presents areas with no data available due to visibility (red and yellow areas) which are in-painted with this technique. In the example of the right, we removed a hard shadow in a building by editing its gradient field and reconstructing the image from it.*

This process can also be used to in-paint small areas in the texture where we do not have information, or even to remove shadows by removing the shadow edge in the gradient field. Examples of in- painting and shadow removal are shown in Figure 5.5.

5.2.4 Discussion

Our formulation of the 'texture reconstruction from multiple views' problem offers several advantages over previous work. Compared to blending techniques, our labelling is unique, avoiding blurring and ghosting effects due to the combination of different images. Our method is easily scalable to a large number of images by performing the optimisation at a lower resolution and then sampling the texture with the resulting labelling at full resolution. Our method is also independent of the mesh resolution and can therefore be applied to highly detailed meshes at the same cost. The gradient based fusion makes a correction to the whole texture, not only at the seams, producing a texture with a unified appearance. The Poisson reintegration phase can be used to repair untextured areas and to remove shadows. Finally, since the optimisation is performed in texture space, the technique can directly take advantage of new parameterisation and atlas packing algorithms, new blending techniques, new solutions for MRF based optimisations, and interactive methods.

5.3 Transferring to a Deformed Space

Once the texture map is reconstructed, we can use it as a diffuse reference to infer albedo and meso-structure by transferring these from exemplars. In the general case, exemplars and texture maps are represented in different domains. Exemplars are photographed using a fronto-parallel view, and are assumed to be globally flat surfaces. The texture or parametric domain introduces an arbitrary deformation per triangle, including rotation, translation, scaling, and warping, with respect to the exemplar domain.

The next section analyses the effect of these transformations in both transferring processes presented in previous chapters – Histogram Matching and Transfer by Analogy – and the solution adopted to perform the transfer in the deformed texture space.

5.3.1 Transformation Texture map - Exemplar

The texture map and the exemplars may have different global orientations, different global scales, and, in the general case, a different mapping per triangle, which means different local piecewise transformations.

Surface Tangential Field

In order to find the relationship between the exemplar space and the texture map space, we fix the horizon as a global orientation for both exemplar and texture map. This means that the exemplar is captured setting the x axis parallel to the horizon. Also that the texture in the complete model is present in the same horizontal direction. For example, for a wall with a texture containing rectangular bricks, we would consider that the long side of the brick is parallel to the x axis in the exemplar, and that is parallel to the plane X, Z in the full model coordinate system. This implies defining a tangential field that traverses the surface in the horizontal direction parallel to the horizon and in the vertical direction parallel to the *up vector*.

Since the geometric model is not necessarily aligned with the vertical axis Y, we first have to define an *up vector*, $\vec{up}$. The user identifies the vertical direction directly in texture space, by viewing the texture as a reference and choosing a straight line where he/she is confident that the texture is not deformed, and marking two texels that establish the *up vector*. Then we can define a tangential field at every texel of the texture map by evaluating the normal vector $\vec{n}$ at every texel and finding the vector orthogonal to both normal and up vectors $\vec{t} = \vec{up} \times \vec{n}$. This tangential field defines the

global orientation in the texture domain. Note that if the exemplar is not a flat surface oriented with the horizon, we can define a tangential field following the same process.

Spatial Transformation

Having exemplar and texture map domains oriented in the same direction, we find the mapping between them. This problem was studied previously for synthesising texture on the parametric domain [LH06, ELS08]. We follow their derivation.
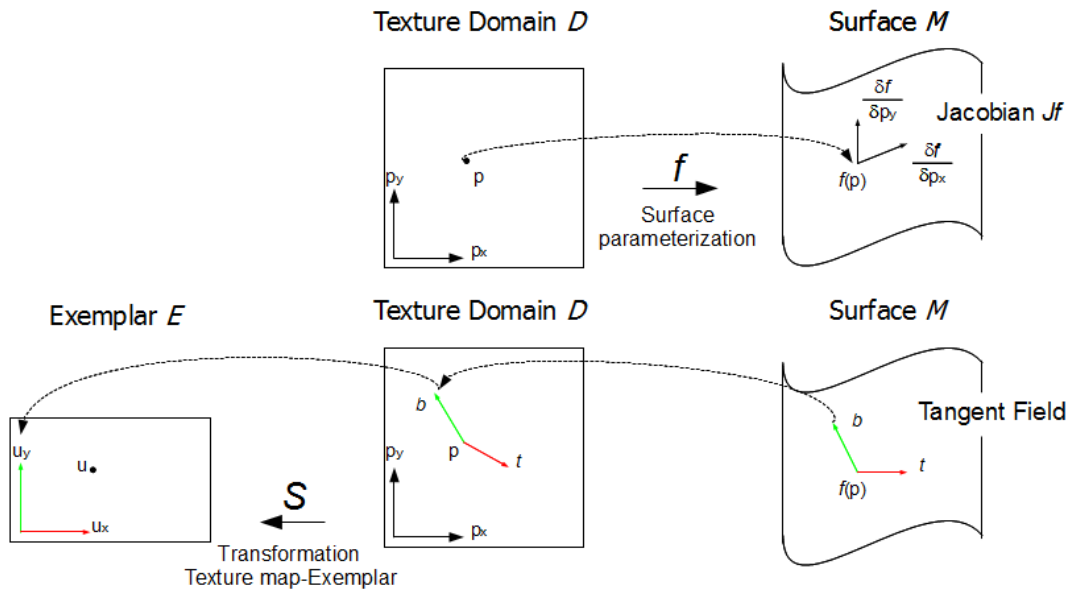
Figure 5.6: *We find the transformation S between the Exemplar Space E and the Texture Domain D to understand how to use our transfer techniques in the Texture Domain from Exemplars.*

Following the Figure 5.6, the surface parameterisation f maps every point in the texture domain D into the surface M. The Jacobian J_f measures the variation of position in the geometry when we vary the position in the texture map. The parameterisation is piecewise linear, so the Jacobian J_f is piecewise constant within each triangle, but we can evaluate it at every texel varying unit vectors in the texture domain D.

Assuming the exemplar to be already oriented, the unitary matrix in equation 5.3 defines its variation, we can define the relationship between the tangent frame and the exemplar as

$$(t\, b) = J_f J^{-1} I \tag{5.3}$$

where J_f is the 3×2 Jacobian of the surface parameterisation $f : D \to M$, and J is the desired 2×2 Jacobian for the relationship between the texture map and the exemplar $S : D \to E$ (see figure 5.6). $(t\ b)$ is the tangential field defined over the surface M, as explained in the previous section. From equation 5.3 we derive

$$J = (t\ b)^{+} J_f = \left((t\ b)^{T} (t\ b) \right)^{-1} (t\ b)^{T} J_f \qquad (5.4)$$

where "+" denotes matrix pseudo-inverse. If $(t\ b)$ is orthonormal, then $(t\ b)^{+} = (t\ b)^{T}$. We construct $(t\ b)$ to be orthonormal, therefore $J = (t\ b)^{T} J_f$. J defines the relationship between the exemplar domain E and the texture map domain D. This relationship encodes local transformations as well as orientation, and is defined at every texel in the texture map.

(a) Reconstructed Texture. (b) Synthesised lines following tangential field. (c) Textured Model.
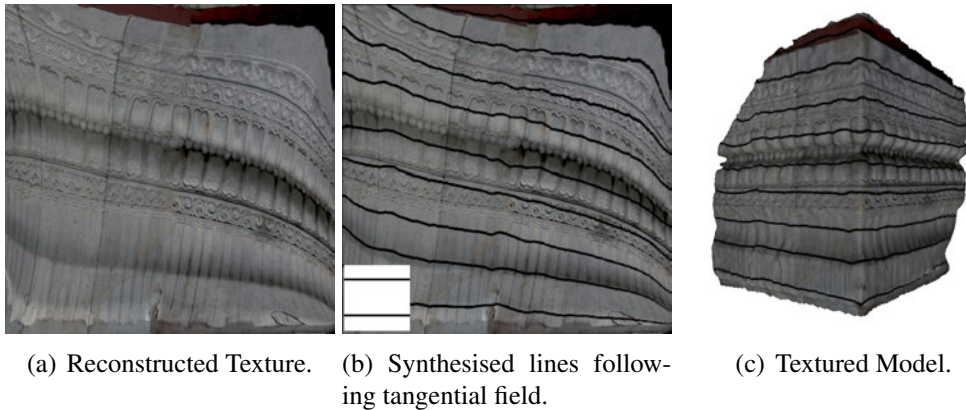
Figure 5.7: *The tangential field defines the vertical and horizontal orientation over the surface of the model. We can use this field to synthesise texture using a simple two lines exemplar. The result is a texture that follows the horizontal and vertical directions of the model.*

In Figure 5.7 we show how we used the Jacobian field to synthesise a texture consisting of parallel lines from a small exemplar (in the middle). The resulting texture is deformed in the same way as the reconstructed texture (left) following the horizontal direction in the geometric model (right). This places exemplar and Texture map in the same space, so they can be compared.

In the next section we use this to analyse the effect of the parameterisation on the transfer process, and also to extend the Transfer by Analogy method to non-planar surfaces.

5.3.2 Histogram Matching

Histograms are used to analyse and index images, and, in our case, to characterise materials. It is important to understand how differences between exemplar and texture map domains influence the histograms. Hadjidemetriou et al. studied more fundamentally the influence of local transformations on the histogram [HGN01]. Here we analyse how changes in the histogram influence the transfer process.

The first thing to note is that the Histogram Matching process uses the Cumulative Distribution Function, which is normalised and therefore invariant to scale. Consequently scaling of the histogram does not influence the transfer process.

Hadjidemetriou et al. derived analytically the complete class of local image transformations that preserve the histogram or simply scale its magnitude. They presented the transformations as vector fields acting on the image and deduced that when these fields are *Hamiltonian*, the histogram is simply scaled. Some simple examples of Hamiltonian transformations are rotation, translation, shear, and stretch [HGN01].

We can consider the exemplar-texture mapping presented before as an image transformation of the exemplar view. Hence, we can define the mapping as Hamiltonian if the transformation of the exemplar view is Hamiltonian. For example, an isometric mapping would represent just rotation and translation with respect to the exemplar – both Hamiltonian transformations – therefore the histogram is preserved and the histogram matching approach is valid. Using this methodology, whether the histogram matching is valid for a given mapping depends on the mapping being Hamiltonian.

Hadjidemetriou et al. formulated and verified the following theorem:

"**Theorem 2**. *A family of transformations T_t which arises as the solution of a vector field X scales the histograms of all the images if and only if the vector field has constant divergence for all t. The scale factors are the determinants of the Jacobian of the transformations at any point.*"

This establishes that if the vector field defined by the mapping has constant divergence, then the histogram is simply scaled and therefore the mapping is Hamiltonian. Consequently, the Jacobian field of the mapping J requires a constant determinant at every point to be Hamiltonian. This condition depends totally on the parameterisation, and therefore, only certain types of mapping are theoretically valid for this approach.

Due to the nature of the data however, we can add some cases where, even if the divergence is not constant, the Histogram Matching is still valid. The basic assumption of the transfer process is that for a given material, the statistics are the same locally (exemplar) and globally (texture map). That means that if the mapping is piecewise

constant in divergence for regions big enough to be considered samples of the material
– i.e. statistically equivalent to the exemplar – then the transfer process is still valid,
because the different scaling factors are multiplying the same histogram, so the final
histogram is simply scaled by the sum of the scaling factors.

It is worth recalling that the transfer process is performed on segments and therefore these requirements are not for the complete texture map but for the individual
segments.

In our experiments, we use parameterisation techniques that create close to conformal mappings, so distortions are minimal. Also, we restrict the possible scaling of the
triangles to 10%. The resulting parameterisations do not produce visible effects in the
final model. However, a more detailed study of the acceptable levels of distortion is
suggested as future work.

5.3.3 Transfer by Analogy

The *Transfer by Analogy* process matches the most similar patches in both images.
Clearly, in order to compare two images, both must be in the same domain. The
exemplar-texture mapping is defined at every point by J, allowing us to transform
one of the domains into the other. The transfer process is performed in a multi-scale
fashion (see section 3.4), therefore, we have to take into account this deformation when
constructing the different levels of the *Laplacian Pyramid*.

Deformed Laplacian Pyramid

The Laplacian pyramid is constructed by blurring and under-sampling the image by
a factor of two. Blurring uses neighbouring information so this process requires to
take into account the deformations produced by the parameterisation. We access the
neighbour texels normally but calculate the distance to the centre of the kernel and the
value of the Gaussian kernel, multiplying the displacement vector by the Jacobian J,
which creates a deformed Gaussian kernel according to the parameterisation. Then, we
under-sample normally. This way we take into account the Jacobian in the calculation
of the Laplacian Pyramid, which produces a correct multi-scale representation of the
texture map. This is presented in more detail in the work of Eisenacher et al. for texture
synthesis on a deformed space [ELS08].

Deformed Patch Matching

The patch matching process performed during the Transfer by Analogy process accesses neighbouring pixels in 7×7 patches. We use the Jacobian J, calculated previously, to modify the neighbourhood of the patch in order to access equivalent pixels in both the texture map and exemplar. The texture-map space has a lower sampling rate than the exemplar and it is also deformed by the parameterisation. Because of this, we access a regular patch in the texture-map and then sample the exemplar according to the mapping J. The deformed patch is computed by accessing the central point and then computing the deformed displacement from the central point by multiplying the displacement by the Jacobian at this position, as shown in Figure 5.8.
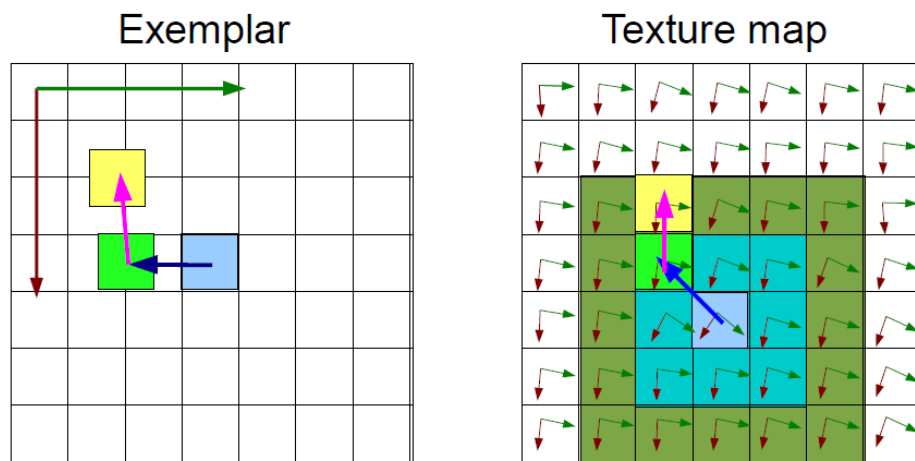
Figure 5.8: *In the texture map (right), we access the neighbourhood of the pixel normally. When calculating the equivalent pixel in the exemplar, we multiply the displacement vector recursively by the Jacobian (green and red arrows). We perform bilinear interpolation for the access coordinates in the exemplar.*

As discussed in chapter 3, the scale between the exemplar and the texture map needs to be specified by the user. This can be included in the Jacobian J by simply multiplying it by the corresponding scalar.

Using this deformed patch matching we are able to perform the transfer process to the texture space. This effectively extends our TbA method to work over arbitrary surfaces, and therefore to complete models.

5.4 Multi-source Geometry Fusion

The last step in the transfer of the material appearance to full models is to combine gross-scale and meso-scale geometries.

Both geometry sources, the gross-scale model recovered from multi-view images, and the meso-scale detail transferred from the exemplars to the texture map, are partially inaccurate. The gross-scale geometry has a limited resolution and suffers the limitations of image-based reconstruction techniques. The meso-scale geometry is an approximation based on shading information encoded in the texture. The reconstructed texture contains artifacts and in some cases violates some of the assumptions, introducing geometric artifacts. This section discusses the nature of these inaccuracies and errors in both sources of geometry, and presents a frequency based combination that improves the result compared to a naive combination. A similar approach was presented in [NRDR05] for merging normal maps and low resolution models. The sources of their data and noise are different to ours, but they arrived at similar conclusions and solutions.

Geometry Images

The mesh and the meso-scale geometry are related by the parameterisation, which associates the texture coordinates and therefore the meso-scale map with every vertex in the model. The mesh and meso-scale maps have different resolutions, so we first have to set both geometries to the same level of detail. We evaluate the geometry of the gross-scale model at every texel, storing its 3D coordinates in the RGB channels of a geometry image [GGH02]. The geometric coordinates for a pixel p inside a triangle t are calculated using the barycentric coordinates of p in t and then multiplied by the 3D coordinates of the vertices of t. The resulting representation of the geometry (shown in Figure 5.9 (left)) allows us to compute its normal map (Figure 5.9 (right)) and directly add the geometric detail computed from the transfer method.

In the next section we show how to combine gross-scale and the meso-scale models filtering out artifacts from both models using a frequency based approach.

5.4.1 Frequency Based Geometry Fusion

Having both geometries at the same resolution and in the same domain, the naive approach to combining gross-scale and meso-scale geometry is to add the depth recovered by the shape from shading algorithm in the gross-scale normal direction. The
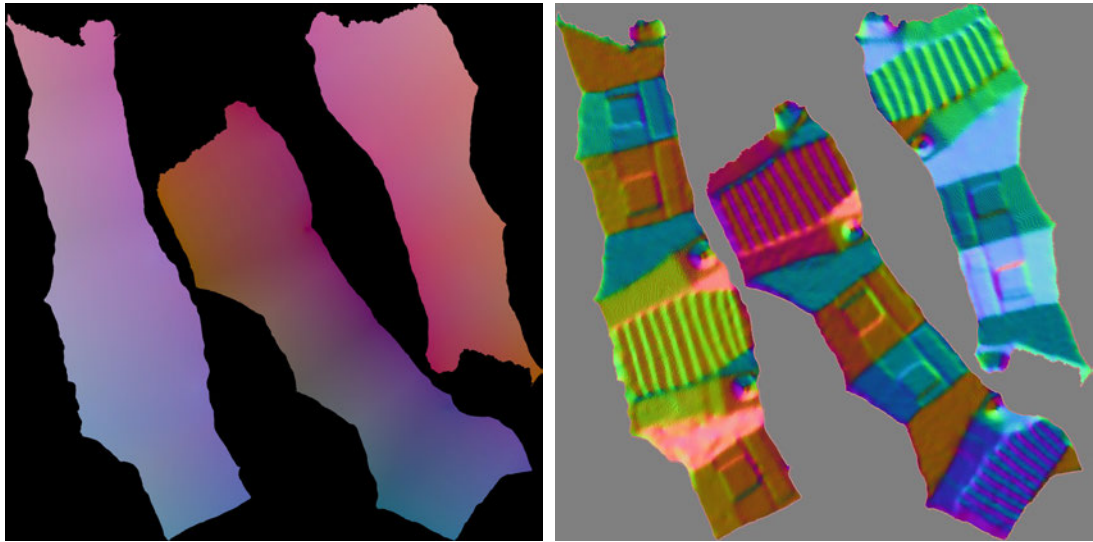
Figure 5.9: *The geometry image (left) places the geometry in the same space as the texture map, allowing for direct fusion of the gross-scale and meso-scale models. The normal map (right) is used to define the direction in which the meso-scale is added to the gross-scale model.*

scale factor between both geometries is related to the distance from the camera to the model, which is difficult to recover automatically without extra measurements or references, since the gross-scale geometry is reconstructed up to a scale factor. This scale factor is specified manually.

A more sophisticated method to combine both sources of geometry, by looking at the errors and noise that both models contain, is proposed here. In general, gross-scale geometry contains high-frequency noise like the bumps visible in figure 5.10(a). These errors are caused by noise in the original point cloud, resolution limitations in the images, errors introduced by the normal estimation and by the surface meshing method. Also, high-frequency geometry like edges and geometric detail are not well captured in general for similar reasons. Therefore, high-frequency information in the gross-scale model is often unreliable. Running a low-pass filter over the geometry removes the high-frequency geometric features and effectively smooths the surface. We can run this filter over the geometry image, accounting for the deformations due to the parameterisation, in the same way as presented for the Deformed Laplacian Pyramid in section 5.3.3. Alternatively, we can use similar smoothing techniques in the geometric space, available in most software packages.

On the other hand, in our approach, meso-scale geometry is used to add the detail

(a) Untouched Gross-scale

(b) Gross-scale model after filtering high frequencies

(c) Real image of the building

(d) Untouched Normal Map

(e) Normal Map after filtering low frequencies

(f) Final model combining both sources of geometry
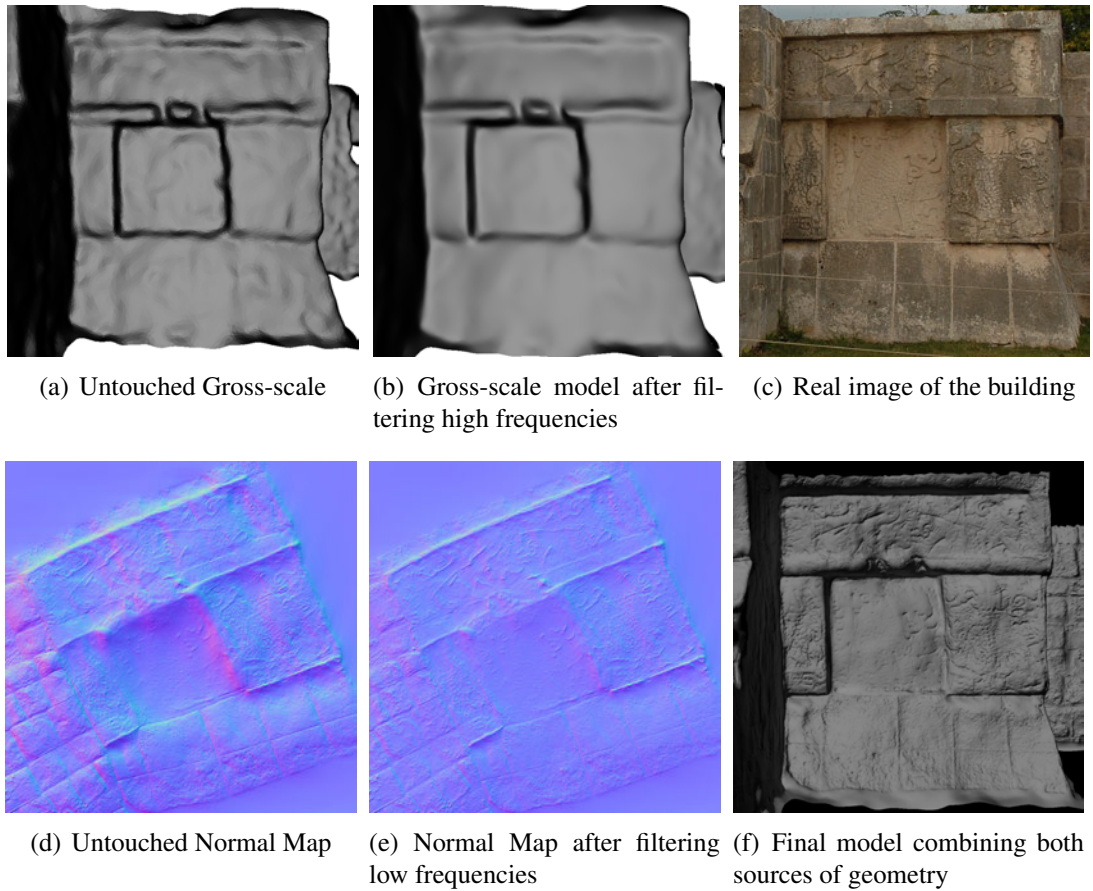
Figure 5.10: *The real image of the building (c) shows how the structure consists of flat panels decorated with carvings. Filtering the high frequencies in the gross-scale model(a - b) and the low frequencies in the meso-scale model(d - e) eliminates the bumps and inaccurate detail, resulting in the final model (f).*

missing in the gross-scale model. This geometry is reconstructed from a shading image and contains unreliable information of a different nature. The shading in a large structure not only contains the effects of texture self-shadowing, that are important for the meso-scale geometry, but also others due to global structure. These shading effects lead to geometry that is not necessary correct and, in any case, is not the geometry that the meso-scale model should contain. In the perception literature, such as Land's *Retinex Theory* [LM71], high-frequency variation is usually attributed to texture, and low-frequency variations are associated with illumination. In our scenario, some of the variations due to illumination are essentially produced by gross-scale features, that are already recovered in the gross-scale model.

Texture can also contain low-frequency information and this implies low-frequency

geometry. This texture information is normally misinterpreted in the transfer process as shading information, when it is actually colour information. Thus low-frequency geometry may recover large geometric features, due to gross-scale shading, or wrong features due to large untextured surfaces shown in Figure 5.10(c). These large features, if present, should have been already recovered by the gross-scale geometry; it is fair to assume that low-frequency geometry at the meso-structure should be treated as either erroneous or redundant, and in any case unreliable. We run a high-pass filter over the shading image, prior to computing the depth map. This retains the high frequency detail in the meso-scale model and removes the artifacts due to global shading or texture misinterpretation (see Figure 5.10(d)).

Figure 5.10(f) shows the final combined model. The combination of both sources of geometry produces a realistic result, approximating correctly the shape and meso-structure of the real building (Figure 5.10(c)).

5.5 Discussion and Future Work

We have presented solutions for the three remaining problems in our pipeline: multi-view texture reconstruction, material transfer to parametric space, and combination of gross-scale and meso-scale models.

Our implemented system for texture reconstruction is able to produce high resolution textures with low levels of distortion and minimise the artifacts, with a quality level appropriate for our application. The reader can see more complete results in chapter 6. Our novel formulation by adapting the ideas of image mosaicing, are shown to work well, producing seamless texture maps.

In this chapter, we also presented an extension to our Transfer by Analogy algorithm to work over the texture domain, solving the remaining limitations of the transfer technique. We also analysed theoretically the effect of the parameterisation in the Histogram Matching process.

Finally, our fusion method for geometry combination deals with the errors introduced by both sources of geometry, reducing artifacts and noise that appear during the process, and improving the quality of the final model. We experimented with the bilateral filter, which preserves edges better than simple low-pass filters, for reducing the artifacts of the gross-scale model. Preliminary results were promising, and we find this an interesting area for further research.

In the next chapter we present the resulting models after applying the complete

process presented in this thesis.

Evaluation of the Complete Reconstruction System

I<small>N</small> *this chapter, we review the complete building reconstruction system, including the techniques presented during this thesis. We show and discuss the results of three architectural structures of historical interest: Clifford's Tower in York, U.K. and the Eagles and Jaguars Platform in Chichen Itza, Mexico; and a façade of Church of the Holy Name in Manchester, UK.*

6.1   System Overview

We first revisit our system pipeline, previously presented in chapter 2, in order to review the complete reconstruction process. The overview of our 3D reconstruction pipeline is shown again in Figure 6.1, where orange boxes are techniques developed and implemented for this thesis.

Our 3D reconstruction system obtains a quasi-dense point cloud, camera parameters for each photograph, and a corresponding surface mesh for the photographed structure. For large structures, requiring a large number of images (more than 20), the model is recovered in several sequences, and partial point clouds are filtered and aligned. The resulting mesh is cleaned by cropping the interesting areas. We refer to the 3D model recovered from this part of the system as *gross-scale geometry*, and in the context of our pipeline this represents the global 3D shape of the acquired façade.

To complement this global shape information with surface detail, we first begin by capturing accessible samples (at close range) of the different textured materials present in the façade, by applying surface depth hallucination. This yields *exemplars* for each material consisting of both albedo and meso-structure. Using the reconstructed model and camera parameters we create a high-resolution texture mosaic for the complete
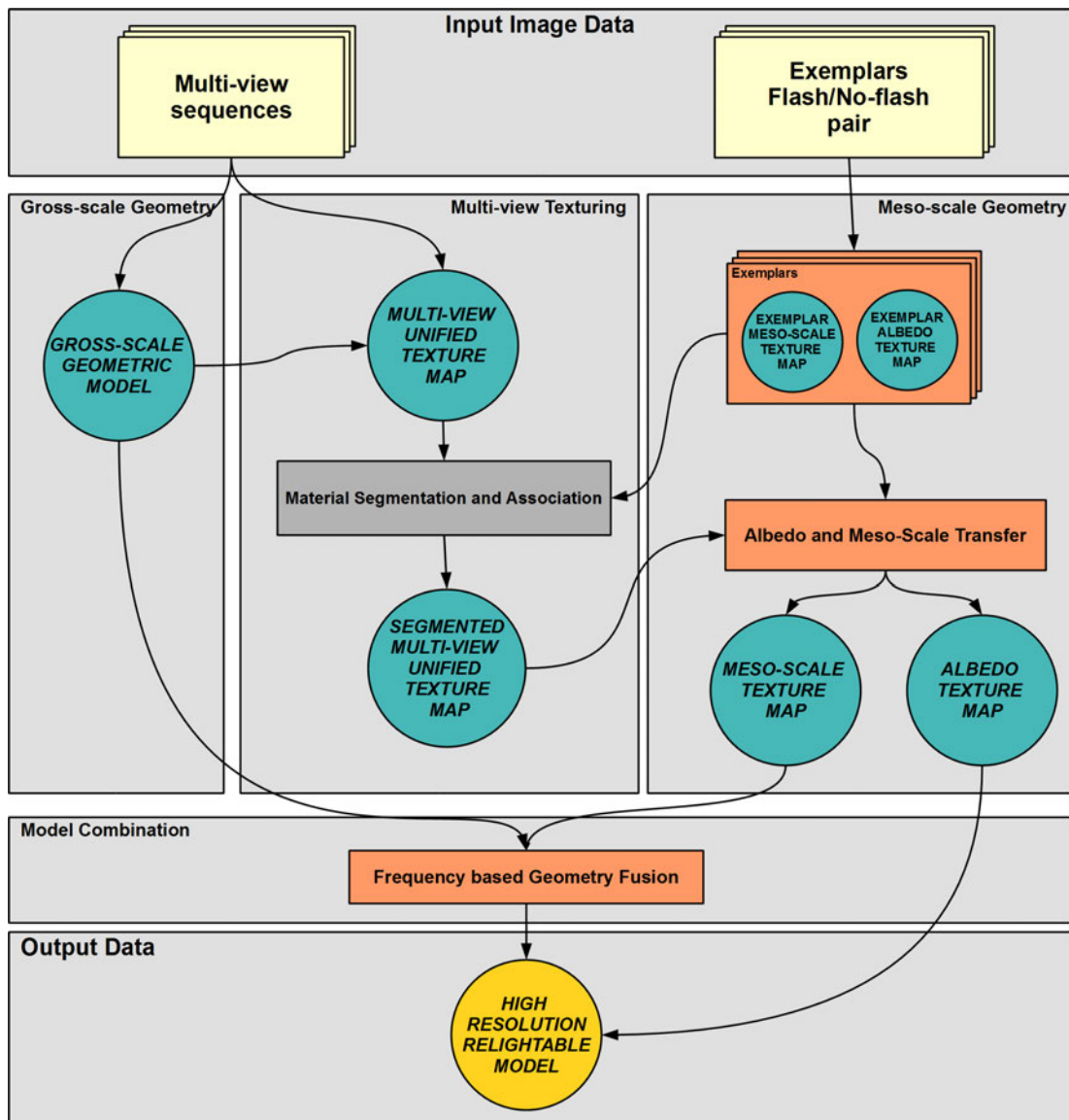
Figure 6.1: *Schematic of our Mixed Image-based Modelling System.*

model from the multi-view image sequence by selecting the best view for each texel as described in chapter 5.

Next we segment the texture mosaic image and assign materials to appropriate exemplars. This segmentation can be performed manually, using commercial software, using our implemented graph-cut based segmentation, or, in the case of Transfer by Analogy, can be performed automatically by using the extension to multiple materials presented in chapter 4.

We then transfer albedo and shading from the exemplars to the segmented texture using Histogram Matching or Transfer by Analogy. In the case of Transfer by Analogy,

we require the user to define an up vector in the texture map, in order to derive a tangential field and recover the transformation between exemplar and texture domain, as described in the second part of chapter 5. Finally, we estimate a depth map from the transferred shading image, after filtering out the low frequencies, which can either be fused with the gross-scale geometry or rendered as a bump map. The fine detail in the complete 3D model results in a visually rich and faithful appearance of the original surfaces.

After data capture our entire process, with the exception of segmentation, is fully automatic. The final models acquired with our method contain highly detailed geometry, and approximate albedo, providing rich appearance under novel lighting conditions. This improves the results of both gross-scale geometry acquisition, and surface depth hallucination, used in isolation. Our method is able to capture interesting surface features that would be labour intensive to generate manually. The breakdown of effort for the models included in this chapter is presented in Table 6.1. The times presented in this table are estimates and their objective is to provide a sense of time consumed by each part of the process and a relative scale between them. The complete process took around one day of work including capture and processing, the segmentation process being the most time consuming.

Table 6.1: *Breakdown of effort. $York_1$ and $York_2$ = Respectively, Exterior and Interior Façades of Clifford's Tower, $E\&J$ = Eagles and Jaguars Platform, $Church$ = Church of the Holy Name. (A) = Automatic. (I) = Interactive. Processing over $7K \times 7K$ pixels Textures.*

	$York_1$	$York_2$	$E\&J$	$Church$
N^o Images Gross-scale	33	7	19	5
N^o Exemplars	5	6	1	3
N^o Segments	16	17	18	4
Capture Gross-scale	30min	10min	20min	10min
Capture Exemplars	1h	1h	10min	30min
Process Gross-scale(A)	1h 30min	30min	1h	25min
Model Cleaning(I)	1h	20min	1h	15min
Texturing(A)	15min	15min	15min	15min
Segmentation(I)	3h	3h	3h	1h
Transfer(A)	15min	15min	15min	10min
Depth Estimation(A)	1min	1min	1min	1min
TOTAL Reconstruction Time	7h 15min	5h 15min	6h	2h 45min

6.2 The Platform of the Eagles and the Jaguars

The Platform of the Eagles and the Jaguars is an elaborately carved platform located on the central plain of Chichen Itza, a Mayan archaeological site, in Mexico. It presents an interesting global structure and lots of surface carvings at the meso-scale level, which makes it ideal to test our system.

Captured Data

Figure 6.2: *Ten of the 19 photographs in the wide-baseline sequence.*

During the data capture process, we collected a set of photographs around the platform every 10 to 15 degrees, covering two thirds of the platform with 19 views. This is a narrower baseline than is normally required due to the structure of the platform, which contains concave angles, requiring more images to cover the façade. A subset of the sequence is presented in Figure 6.2. Besides the wide-baseline we took a flash/no-flash pair of photographsfrom which we derive an exemplar that contains a representative panel of the platform (see Figure 6.3), which we can use both for transfer and comparison.

Figure 6.3: *Resulting exemplar for the Eagles and Jaguars reconstruction model using Surface Depth Hallucination. From left to right, resulting albedo map, shading map, and final relit model.*

This data was captured by Mashhuda Glencross during a trip to Mexico. She had no personal experience with the wide-baseline reconstruction system, but just a general idea of how to capture the images. This conveys informally the easiness of the capture process. Unfortunately, the exemplars were captured before the idea of transferring materials was developed. For this reason, we do not have data for other types of texture. However, the results using only one exemplar are compelling and visualy rich.

Texture Map

Using our texture reconstruction system, we produced a high-resolution texture map. With the 19 views available, we are be able to perform the optimisation at a 1500×1500 resolution and then super-sample to the final 7000×7000 texels. We show the complete texture map in Figure 6.4.

Figure 6.4: *Eagles and Jaguars $7K \times 7K$ texels texture map reconstructed from 19 views. The red square indicates the area corresponding a particular area that we will discuss further in the next sections.*

The red square indicates the zone corresponding to the same area recovered by the exemplar in Figure 6.3. Before showing the complete model under novel lighting conditions, we will focus on this marked area, and will discuss the results of the transferred albedo and geometry compared with the depth hallucination method, and with another image-based geometry reconstruction method.

Eagles and Jaguars Panel

We focus on a section of the platform, that we refer to as *panel*, in order to judge the quality of the transferred albedo map, the transferred meso-structure, and the compete model. We chose this particular panel because we have an exemplar of the same section which we used to compare and judge the results.

(a)   (b)   (c)
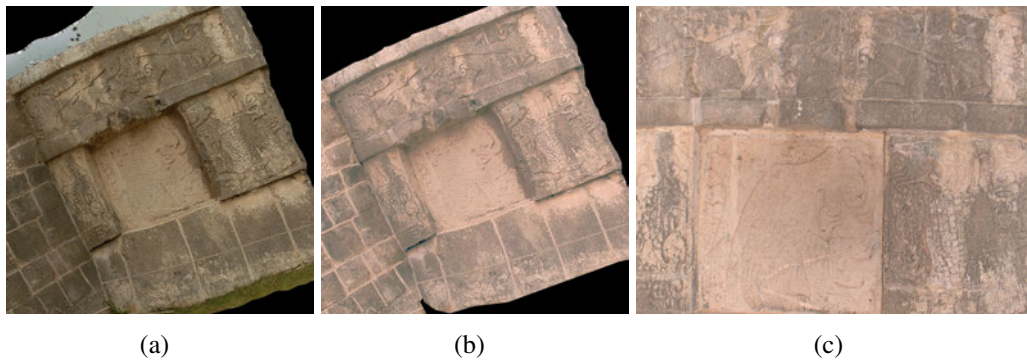
Figure 6.5: *(a) Section of the texture map corresponding to the panel, (b) Transferred albedo using Histogram Matching, (c) Same section of the platform captured using the Depth Hallucination method.*

In Figure 6.5, we show the region of the texture map of the complete model corresponding to the panel (a), the transferred albedo map for this region (b), and the captured exemplar (c). The distortions and change of orientation in (a) and (b) with respect to (c) are due to the parameterisation of the geometry. We appreciate that the appearance of the panel in (a), under the capturing lighting conditions, differs significantly from the reflectance recovered with the depth hallucination method (c). The appearance of the histogram matched albedo (b), approximates the reflectance of the panel consistently to the captured one (c), effectively un-lighting the texture. The transfer method does not suffer due to the small distortions introduced by the parameterisation process, following the behaviour analysed in previous chapters.

Figure 6.6 shows the material transfer to the same area, in this case performed using the transfer by analogy technique. The albedo map has an appropriate colour,

however, it has lost some structure. Some areas have been matched perfectly, finding the corresponding features in the exemplar, and transferring the albedo and shading accordingly. However, some areas are not correctly matched because of two reasons. The first one is that some features are not present in the exemplar. The areas at the left and the bottom of the panel are not contained in the exemplar, and therefore, this method failed to estimate the albedo of these areas. The second reason is the sensitivity of the method to scale and orientation. If these two parameters are not perfectly matched at every point, the method has problems to find the correct feature. In this particular example, the method matches perfectly some features, as can be seen in the transferred shading map in (c), but fails in other regions.

(a) (b) (c)
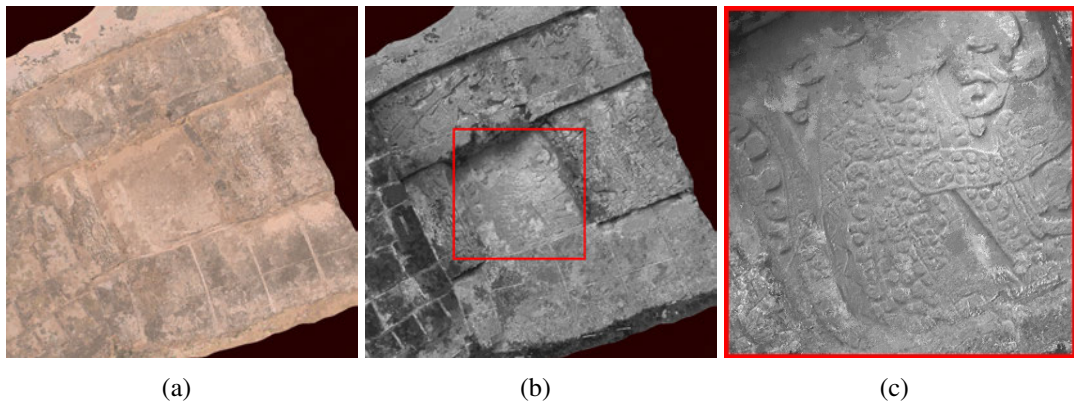
Figure 6.6: *(a) Albedo for the panel transferred using TbA, (b) Shading transferred using TbA, (c) Close up of the middle section of (b).*

The results from the TbA method are less consistent than the ones produced by the Histogram Matching method. However, when the features are matched correctly, as happened in the area shown in Figure 6.6(c), the resulting albedo and shading maps are more accurate.

Geometric Models

Figure 6.7 shows visual comparisons for the geometric models of the panel, reconstructed using different techniques. Our combined model (6.7(a)) is able to recover both global depth and local detail, while the model containing only gross-scale geometry (6.7(d)) fails to recover meso-structure. Comparing with [DTC04] – considered the gold standard in automatic multi-view reconstruction – their model (6.7(c)) provides sharper geometry than our gross-scale model, but introduces artifacts at the fine detail level. Our transfer system provides more detailed and coherent meso-structure.

137

Depth Hallucination (6.7(b)) captures high-resolution local detail, but is not able to model the global structure, presenting little depth difference between the centre of the panel and the rest, due to the flat-surface assumption.

(a)   (b)

(c)   (d)
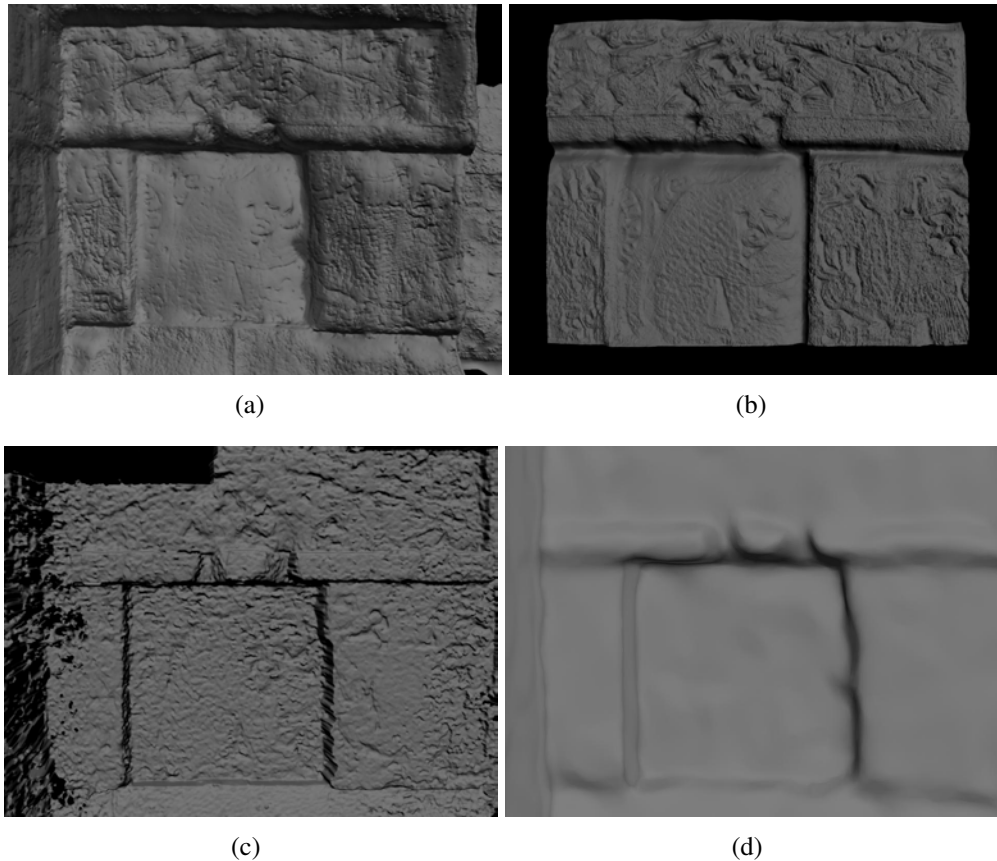
Figure 6.7: *(a) Close-up view of panel reconstructed with our transfer system, (b) using Depth Hallucination [GWJ+08], (c) reconstructed with [DTC04], (d) gross geometry without meso-scale detail.*

It is important to emphasise that (b)(SDH) was created using an optimal view of only this panel. In contrast, (a) was constructed from views of the whole building matched with the exemplar. Some minor differences in detail can be seen, but overall results are quite similar.

In general, we consider that the geometry resulting from our complete process improves the results of gross-scale and meso-scale models separately, and compares favourably with the method of Dick et al. [DTC04]. In particular, the meso-structure improves the lack of sharpness at the edges of the gross-scale model.

Relit Models

Finally, we present renderings of the complete model under novel lighting conditions. We used both physically based sky models, and image-based lighting with an environment map. The renderings are compelling and show a great level of detail in the carvings.

Figure 6.8: *The model shows a great level of compelling detail and a realistic appearance.*

The image in Figure 6.8 shows a rendering of the complete model under cloudy lighting conditions. Looking at the close up view, we can observe self-shadowing effects produced by the detail acquired with our system, for example in the carvings of the panel on the left, or the jaguar head on the right.

To finish with this model, we show in Figure 6.9, the platform rendered under four different lighting conditions: morning (a), late afternoon (b), sunset (c), and night (d). The appearance of the model changes to reflect these conditions and yields results that
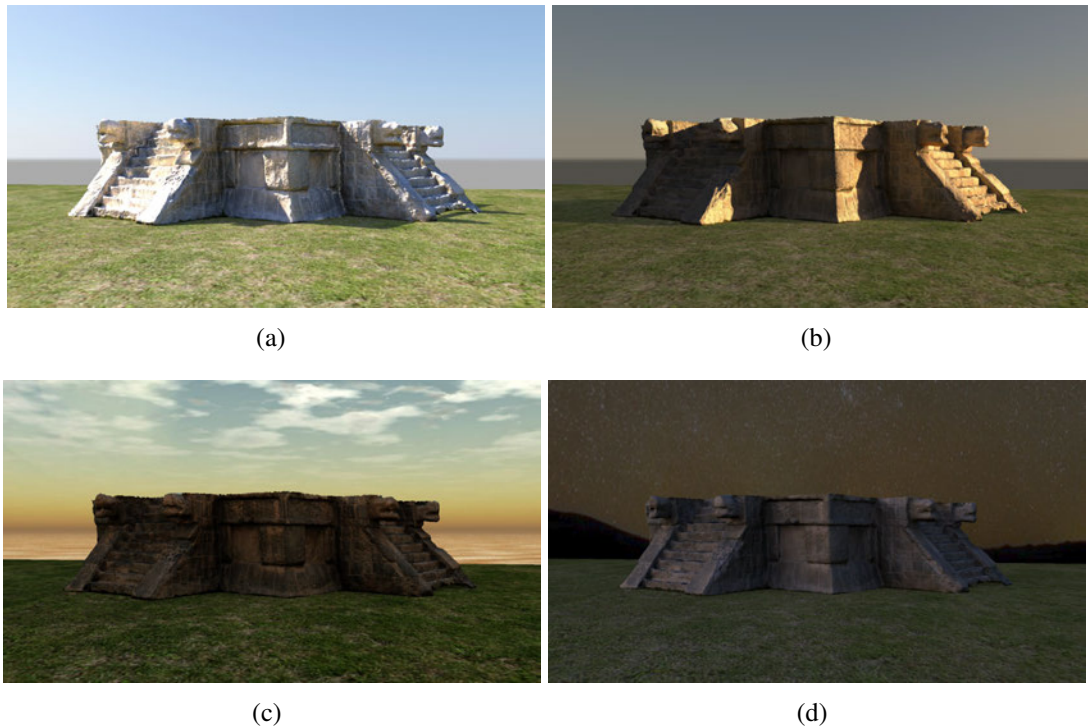
Figure 6.9: *The platform of the Eagles and the Jaguars model rendered under different lighting conditions.*

correspond visually to how we might expect them to look.

6.3 Clifford's Tower

Clifford's Tower is a historic castle in the English city of York. This building has an irregular structure and the stone around it has deteriorated, providing interesting textures and colour variations. We reconstructed the exterior façade and a section of the interior.

Figure 6.10 shows some renderings side by side with images downloaded from the Internet [1] [2] [3] matching the view point. The lighting conditions have not been matched, but the appearance is realistic and plausible. Despite the presence of minor artifacts, for example in the roof and windows, both model and reflectance produce realistic looking results comparable to the photographs.

The synthetic renderings in Figure 6.10 show the appearance of the model under

[1] (b) Source:http://tigg-stock.deviantart.com/art/Castle-stock-10-73841672
[2] (d) Source:http://www.nicolaconforto.com/erasmus/gitaayork.php
[3] (f) Source:http://www.english-heritage.org.uk/daysout/properties/cliffords-tower-york/

Figure 6.10: *Renderings of the reconstructed model using a physical model of the sky (left column) and a set of photographs, downloaded from the Internet, matching the view points.*

direct lighting (a), producing more pronounced self-shadowing effects, and in shadow (b and c), where these effects are less apparent. Again, the appearance of the building under new lighting conditions is consistent with the appearance in the photographs. Surface details like the carving at the entrance are well modelled and consistent with the texture. The importance of these details can be seen in Figure 6.11.

141

Geometric Models

At the scale of the complete building, meso-structure includes inter-stone cracks, windows, and shield carvings above the entrances. Without this detail, the model (left image in Figure 6.11) appears somewhat flat. In contrast, our detail-enhanced model introduces all these aspects of the geometry.

(a)   (b)

(c)   (d)   (e)   (f)
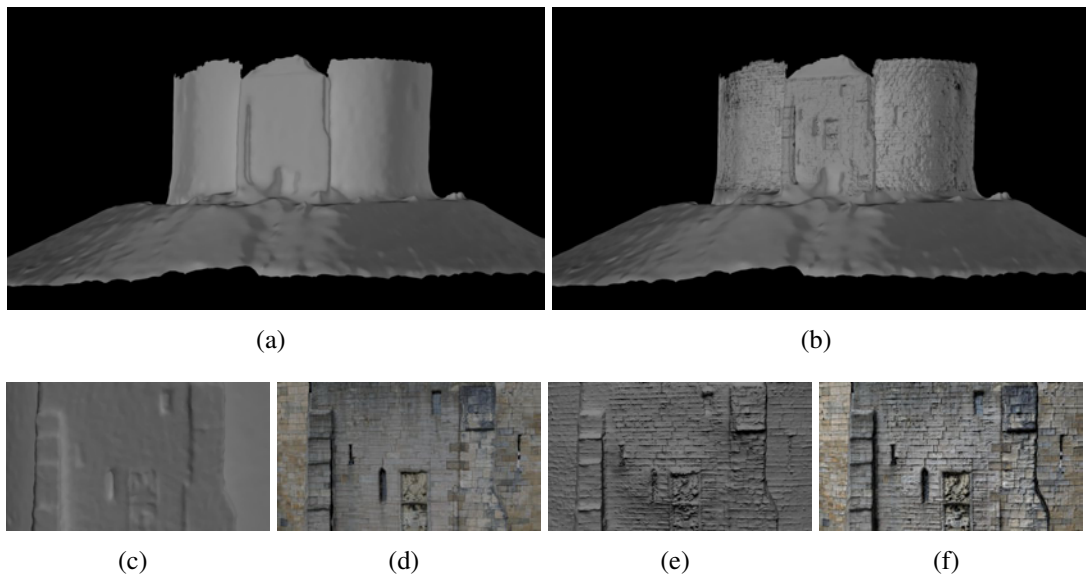
Figure 6.11: *(a, c) Gross-scale model only. (d) Gross-scale model with albedo texture. (b, e) Complete reconstructed model including meso-scale geometry. (f) Complete model including meso-scale geometry and albedo texture*
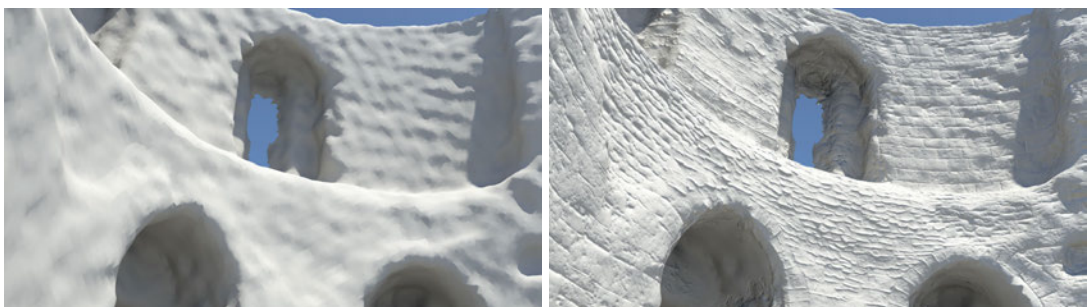
Figure 6.12: *Portion of the interior façade of the Clifford's Tower Model, reconstructed with our system. Again the transferred detail enhances the model, adding important features such as the cracks between stones.*

The close up view of the entrance façade in the second row of Figure 6.11, shows comparisons between the gross-scale and complete models with and without texture.

The meso-scale detail changes the appearance radically when the lighting changes. Without the meso-scale (c), the model has an appearance similar to a surface covered with wall paper independently of the lighting. The meso-structure is important for the visual richness of the render and for the variation of the appearance with the change of light.

Similarly, in the case of the interior façade in Figure 6.12, results show the enhancement of the geometry due to the added geometric detail. The cracks in the stone wall, missing stones, and other holes also mask some of the artifacts due to the gross-scale geometry such as the small bumps in the left image.

Relit Models

Figure 6.13: *Exterior façade of the Clifford's Tower Model, reconstructed with our system, rendered under different view-points and lighting conditions.*

Finally, we present some relit renderings of the exterior and the interior of Clifford's Tower in Figures 6.13 and 6.15, illustrating the quality and realism of the reconstructed models.
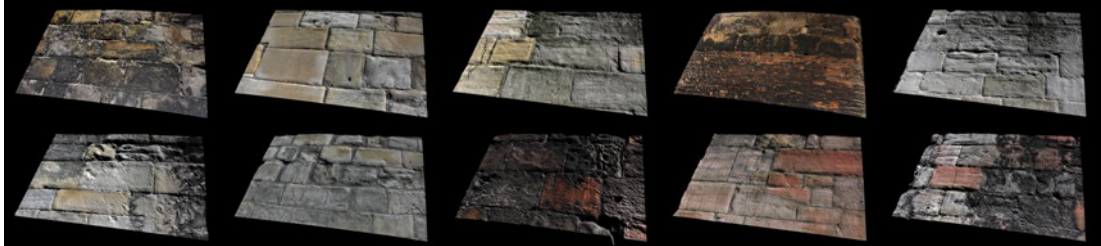
Figure 6.14: *Exemplars of the 10 different materials captured for these reconstructions.*

We can see how the cracks and carving details change the appearance of the model under the different lighting conditions. Despite the minor artifacts on the stairs, and other non-modelled objects (pillars, handrails, etc), due to the limitations of the gross-scale reconstruction system, the final renderings provide realistic results suitable for many types of applications such as video games.

The interior façade exhibits a high variation in the types of material. Figure 6.14 shows the ten exemplars captured for Clifford's Tower. Most of these were captured in the interior. This variation makes segmentation harder and correct assignment of materials more important. The reconstruction of the interior façade (Figure 6.15) shows a comparison with a photograph of the same section of the building. Although the surface detail is nicely recovered, some artifact in the albedo are present. For example, there is an excessively pink area in the middle of the surface. Also, the small stones in the middle of the structure are clearer than the ones in the photograph. This is caused by wrong assignments during the material-exemplar association process, in the case of the small stones due to the inaccessibility of this type of texture. Where the association is correct, the material appearance is well approximated. A more accurate result could be achieved by performing a more exhaustive segmentation.

The models presented in this thesis were created with minimal manual intervention, to illustrate the capabilities of our automatic processing. Other visible inaccuracies in the geometry shown in Figure 6.13, like the window holes or the white area on the top of the model, can also be removed by performing a more detailed cleaning.

Figure 6.15: *Interior façade of the Clifford's Tower Model (top and bottom-right), compared with a photograph of the real façade (bottom-left).*

6.4 Church of the Holy Name

This church is situated on the campus of Manchester University. It has a highly ornamented façade as can be appreciated in Figure 6.16 (right), with columns, carvings, small towers, etc. We took only five photographs of the building in order to reconstruct the gross-scale model, and one exemplar of the type of stone covering the major area

of the façade. The dark type of stone in the higher section, was approximated manually by modifying the captured exemplar. Windows were matched to a black window captured in a different place and doors were textured in black. Windows in the façade are highly specular and we do not recover this material property. The resulting modelled façade is shown in Figure 6.16 (left), rendered under novel view point and lighting.
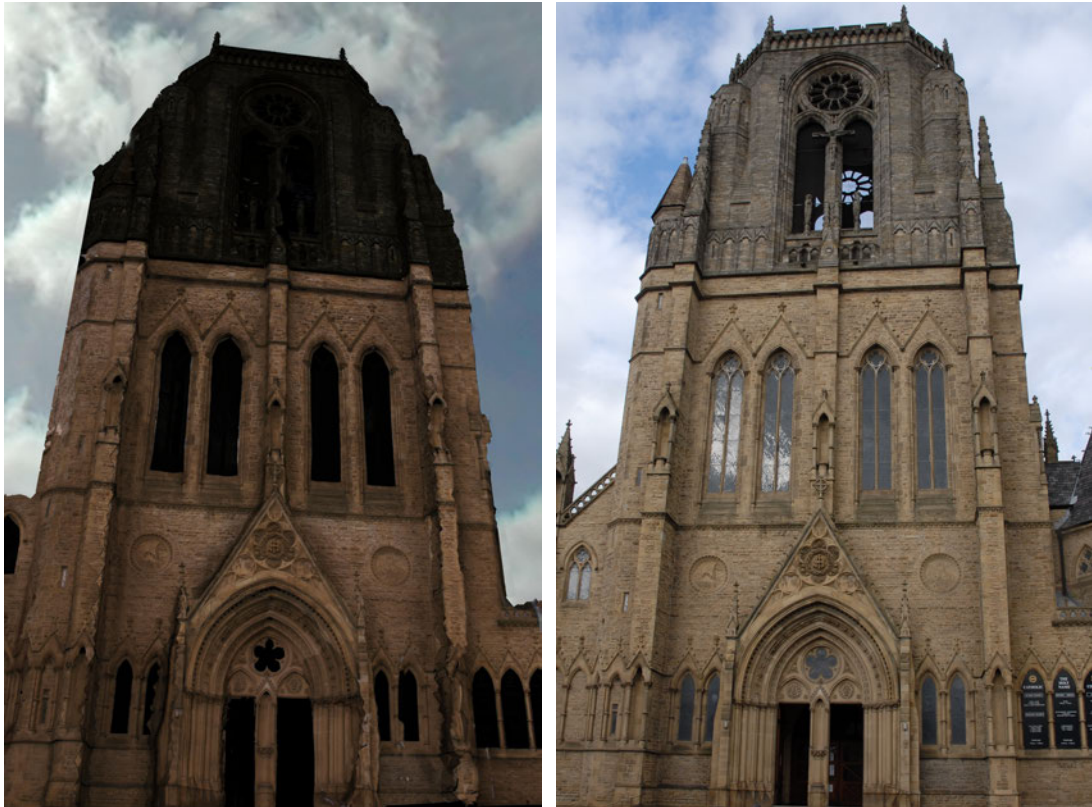
Figure 6.16: *Façade of the Church of the Holy Name. (Left) Reconstructed model under novel lighting and view point. (Right) Photograph.*

Most of the limitations in our system arise from the gross-scale geometry reconstruction. Structures with complex ornaments are challenging to reconstruct from multiple views, requiring very high resolution images, which cannot be processed in our desktop computer. Also, the meshing technique used in our pipeline has difficulty correctly recovering these ornaments. On the other hand, the carvings and other geometric details that can be modelled by our meso-structure acquisition technique, are correctly modelled as can be seen in Figure 6.17.

Note the two circular ornaments on the sides of the façade, and the central symbols at the top of the entrance. This detail geometry is hardly accessible to capture with other techniques and would be labour intensive to model manually. Our method

Figure 6.17: *Close up view of an interesting area of the façade. (Top) Relit model. (Middle) Recovered geometry. (Bottom) Photograph of the same section.*

succeeded to model the geometry and the visual appearance of these ornaments almost fully automatically.

Figure 6.18: *Façade of the Church of the Holy Name rendered under synthetic floodlit.*

Our last rendering shows our model simulating flood lighting at night. The appearance is also consistent under this lighting conditions and interesting features of the texture show up due to the meso-structure modelled with our system.

6.5 Final Considerations

We show, in the images of this chapter, examples of the high-quality models recovered with our reconstruction system. The meso-structure computed by the transfer techniques developed in this thesis, provides enough geometric detail to reproduce self-shadowing effects when relighting. This detail is consistent with the texture information and presents a significant improvement over the gross-scale geometry model alone.

The approximated albedo is transferred correctly to the complete texture map, reproducing the albedo of the exemplars, effectively un-lighting the texture. The final reflectance map contains small artifacts introduced by the texture reconstruction process, but the final result has a good overall quality. Correct association is crucial to obtain an accurate result. When the correct exemplars are inaccessible, we can manually approximate them using image processing software.

These results provide evidence of the effectiveness of our system on providing reflectance and meso-structure estimates for complex outdoor scenes, by using a transfer approach. This provides very plausible models, with very high level of detail, at a very low cost and effort.

Conclusions and Future Work

THIS *concluding chapter reviews the work presented in this thesis, followed by a summary of contributions and suggestions for future work.*

We have described a novel method for constructing models of buildings from photographs. Our method extends existing approaches for geometry acquisition by adding image-based estimation of local depth and albedo. We illustrated this by applying it to capturing both the geometry and detailed appearance of building façades. Although this detail is approximate, it is informed by shape information contained in the statistics of the photographs. The method mitigates the limitations of image-based 3D reconstruction and surface depth hallucination by fusing information captured at different scales. This combination provides relightable 3D models for a range of cultural heritage, visualisation and entertainment applications.

7.1   Summary of the Thesis

In the thesis we addressed the problem of constructing visually faithful models of building façades. Our objective was for reconstructed models to appear as much like the real scenes as possible, under varying viewing and lighting conditions. The main focus in this investigation was on approximating detailed surface meso-structure and albedo, and combining this with a gross-scale geometric model. We envisaged a scenario where a user takes several pictures of a building from different viewpoints to recover the global structure, and sample images of materials at a close range, used to estimate material characteristics. Surface detail and appearance were transferred from the exemplars to the complete model. This approach yields high-quality models, imparting the illusion of measured reflectance.

Image-based multi-view reconstruction systems allow users to recover good low-to-medium resolution models, but struggle to acquire reliable and realistic texture detail. High resolution texture geometry and reflectance are important for plausible relighting and view point change, and can be acquired by using controlled lighting. However, the specific problems of reconstructing outdoor scenes limit these techniques to the power of portable light sources and, therefore, to accessible small areas. Chapter 2 reviews these methods, presented our pipeline, overviewed the gross-scale reconstruction system and our contributions to it, and motivated the necessity of a novel approach.

The thesis addressed the difficulties of previous image-based reconstruction systems, by exploring an exemplar based transfer method to add meso-scale geometry and reflectance to a gross-scale model. Chapter 3 overviewed the Surface Depth Hallucination technique for capturing material exemplars of textured surfaces. Then, the transfer methods Histogram Matching and Transfer by Analogy were described. Emphasis was placed upon the analysis and performance of both techniques, aiming to gain new insights into their transfer capabilities and their application to complete models. Through the evaluation of these techniques, we concluded that both albedo and meso-structure are reasonably well transferred for different types of textures, but the result is data dependant. Histogram Matching provides plausible and structurally consistent material characteristics, but only removes self-shadowing effects completely in certain cases. This particular limitation is partially mitigated during the rendering process and does not affect the plausibility of the result. Transfer by Analogy does not require the global statistics to match, behaving better in shadowed areas at the cost of losing some image coherence. The combination of the two transfer techniques was briefly explored with promising results.

These methods were extended in Chapter 4 to several materials and large surfaces. This extension consists of segmenting the façade into different regions containing only one material and performing the transfer on a one-to-one basis. We discussed a practical solution for interactive material segmentation using graph-cuts within a Markov Random Field framework. Interactive techniques are recommendable, providing a good trade-off between effort and accuracy. This workflow produced plausible estimates for both albedo and geometric detail. Aiming for a completely automatic system, we investigated the application of data-driven descriptors, inspired by the texture synthesis literature, to associate exemplars with materials in the façade. We observed, in our experiments, that colour descriptors of relatively small sizes (7×7 or 9×9), in combination with a robust-to-noise metric, provided the most coherent segmentations.

In the examples tested, the method provided promising results on close to regular textures, but it had some limitations in its robustness and accuracy. Further studies on texture models based on more complex descriptors and statistical models could benefit our algorithm, improving the matching process.

Chapter 5 discussed the requirements of an optimal texture map and presented a novel approach for its acquisition. Our texture reconstruction system follows the workflow of image mosaic creation by using minimisation techniques over a MRF framework. An optimal solution was achieved by minimising seams and number of patches, and including occlusion detection. Poisson reintegration was used to remove any remaining seams and to in-paint small un-textured areas. In the second part of the chapter, we analysed the effects of the deformations caused by the texturing process in the application of our transfer techniques. Histogram Matching requires the texture space to be a Hamiltonian transformation of the material exemplars, which was achieved by creating a close-to-conformal texture mapping. Transfer by Analogy requires deforming the neighbourhood computation according to the transformation between texture map and exemplars, which needs defining a tangential field on the texture space. Once the transfer is performed over the deformed texture space, we finished the reconstruction process by fusing gross-scale and meso-scale geometries, using a frequency based approach. This filters out artifacts from both sources of geometry, creating a detailed approximation of the building geometry.

An evaluation of the complete system was carried out in Chapter 6. We showed realistic reconstructions of two historic buildings that contain typical challenges of outdoor scene reconstruction. Our models compared favourably against reconstructions carried out with other image-based approaches. Side-by-side comparison with photographs of the buildings showed the plausibility of the relit models and illustrated the effectiveness of our approach. Minor artifacts arose at every stage of the reconstruction system, and models cannot yet be regarded as comparable with the level of accuracy of laser-scanned models or other more labour-intensive methods. However, the modest effort required to create our models proves our system as a really useful tool for low-budget projects, casual users, and applications such as video games, where visually compelling appearance is more important than accuracy.

7.2 Summary of Contributions

The application of exemplar-based texture transfer techniques to the problem of recovering meso-scale geometry and visual appearance for outdoor scenes offers a new approach to the inverse rendering problem. This approach can be used in any existing textured model to enhanced its geometry and visual appearance, including cutting-edge research in automatic reconstruction of cities from image collections. The transfer techniques presented in Chapter 3 have shown their effectiveness on automatically approximating material characteristics for a variety of textures with plausible results. In the thesis we have established the characteristics that can and cannot be transferred by these techniques, their limitations, and their relative merits. This study provides important insights for inexperienced users to decide which technique they should use depending on the type of material.

In Chapter 4, we identified the problems involved in applying our transfer techniques to complex surfaces over a change of scale. Histogram Matching was shown to work well by segmenting the texture and associating every segment with an exemplar for transfer. We also analysed the requirements of the texture parameterisation to correctly use the Histogram Matching technique on non-planar surfaces. The results produced by this method provide visually rich reconstructions, and the simplicity of the process allows it to be applied in other pipelines. At the moment, the results of this technique are being assessed by a major US games company, Activision, to reduce the effort dedicated to modelling normal maps and albedo maps for architectural structures.

The Transfer by Analogy method provides the basis for automatically transferring several materials without the necessity of segmentation. The experiments of this thesis showed the potential of the technique and encourage further research. We solved the problems of applying this technique over complex surfaces containing multiple materials, and over changes of scale, orientation, and light intensity. Close to regular textures benefit from this method and produce better results than Histogram Matching. To completely automate these processes is still an open issue, however, the experiments presented in this thesis contribute to the body of knowledge needed to reach this goal.

A new method for texture reconstruction from multi-view images based on graph optimisation was presented in Chapter 5. Our formulation deals with the common issues of texturing from wide-baseline photographs, including optimal camera selection

for artifact reduction and quality maximisation. Our technique is able to produce seamless high-resolution textures and our definition of an indirection map maintains the continuity of the texture map, even when the parameterisation require cutting the mesh into different charts. Although this method is designed to accomplish the requirements of our specific pipeline, it can be applied to any type of model reconstructed from multiple images.

The complete working system developed during this thesis introduces the individual contributions of the different chapters, providing a tool capable of producing realistic models of full building façades, containing high-resolution geometry and plausible approximations of visual appearance, with modest user interaction. As commented in Chapter 6, the system only requires basic guidelines for data capture, providing complete relightable models in less than a day of work.

7.3 Future Work

Automatic reconstruction of urban environments is of large interest for the graphics and vision community. We believe that new reconstruction systems can benefit from our techniques, by enhancing the realism and visual richness of their models, acquiring surface detail and reflectance. To provide such a benefit, we intend to create a publicly available database of exemplars of common materials to be transferred to global surface models. This would also allow individual users to enhance their existing 3D models. In this scenario, providing automatic material associations can be crucial. Our suggestion for future work is the improvement of texture descriptors for both texture transfer and association, based on patch-matching techniques as suggested in Chapters 3 and 4. Inverse texture synthesis, where textures are compacted in small exemplars by analysing the local neighbourhood, can reduce significantly the size of our exemplars, and therefore the execution times, allowing the system to compare the textures against a large database of exemplars.

Models reconstructed from multiple views suffer from lack of sharpness at edges, especially after filtering out high-frequency information. Using texture information as well as edge-preserving filtering methods, such as the bilateral filter, can improve the gross-scale model. On the other hand, using the gross-scale model to detect shadows in the estimated shading map, could improve the surface detail derived from it.

Finally, transfer of different material characteristics such as specular reflectance,

normal information, or physical properties, is of interest for video games and interactive virtual environment.

Perceptually Validated Surface Depth Hallucination

THIS *appendix presents, for the convenience of the reader, the original Surface Depth Hallucination paper published in SIG-GRAPH 2008 [GWJ$^+$08].*

A.1 Abstract

Capturing detailed surface geometry currently requires specialized equipment such as laser range scanners, which despite their high accuracy, leave gaps in the surfaces that must be reconciled with photographic capture for relighting applications. Using only a standard digital camera and a single view, we present a method for recovering models of predominantly diffuse textured surfaces that can be plausibly relit and viewed from any angle under any illumination. Our multiscale shape-from-shading technique uses diffuse-lit/flash-lit image pairs to produce an albedo map and textured height field. Using two lighting conditions enables us to subtract one from the other to estimate albedo. In the absence of a flash-lit image of a surface for which we already have a similar exemplar pair, we approximate both albedo and diffuse shading images using histogram matching. Our depth estimation is based on local visibility. Unlike other depth-from-shading approaches, all operations are performed on the diffuse shading image in image space, and we impose no constant albedo restrictions. An experimental validation shows our method works for a broad range of textured surfaces, and viewers are frequently unable to identify our results as synthetic in a randomized presentation. Furthermore, in side-by-side comparisons, subjects found a rendering of our depth map equally plausible to one generated from a laser range scan. We see this method as a significant advance in acquiring surface detail for texturing using a standard digital camera, with applications in architecture, archaeological reconstruction, games and special effects.

Figure A.1: *The left image is a photograph of a Mayan carving under diffuse lighting, which was combined with a similar flash photo to derive a height field and albedo map for this surface. The middle image uses the derived model to render the same view with novel lighting. The rendering on the right shows an oblique close-up with a second novel lighting condition and added specularity.*

A.2 Introduction

Textured surfaces such as brick, stone, wood and many other building materials have local variations in their surface *meso-structure*. Shading variations due to self-shadowing provide important perceptual cues necessary to convey a correct impression of shape. An interesting question, however, is how accurate does surface meso-structure need to be for shape and corresponding shading to appear plausible? This is an important question as our objective is to produce synthetically relit results that are perceptually difficult to distinguish from photographs. (See Figure A.1.) In this paper we show that an approximate representation of the real surface (depth + albedo map) may be used to relight predominantly diffuse textured surfaces in a visually plausible manner. To this aim, we introduce a practical method to recover approximate surface texture information from a single viewpoint. From a 2D picture, we infer surface depth where it is not fully divulged in the image. We call this *depth hallucination*.

Representing surface detail is useful to increase the visual realism in a range of application areas, especially architectural reconstructions. In particular, accurately assessing the effect of new buildings on lighting requires modeling of gross 3D geometry, meso-structure, and *albedo* (equivalent to diffuse reflectance), so that simulations of appearance at different times of the day are possible. Our method is aimed primarily at the materials recovery part of such architectural reconstructions. We aim to acquire surface meso-structure so that it may be combined with gross 3D geometry (obtained using another method) to convey the appearance of visually realistic surface detail. However, this technique is equally applicable to recovering and representing surface detail for use in graphically rich games and movies. Currently, our method is being used to recover depth maps at Chichén Itzá (Mexico), for the production of a

dome-projected movie.

Our main contribution is a novel, experimentally validated shape-from-shading method, which takes diffuse-lit/flash-lit image pairs and produces a plausible textured height field that can be viewed from any angle under any lighting. In the absence of a flash-lit image, we apply histogram matching against a visually similar texture for which we have recovered a model from captured pairs. This practical optimization simplifies the capture requirements for large surfaces composed of the same material but containing significant meso-structure variation. Since our goal is to recover enough surface detail for plausible relighting, accuracy requirements are purely perceptual and are evaluated based on the final imagery. To date, no published method for recovering and relighting textured height fields has been validated against equivalent photographs. Our experimental studies demonstrate that participants cannot reliably identify our re-lit images as synthetic, and more importantly that they believe these to be as plausible as geometrically correct laser-scanned reconstructions.

A.3   Previous Work

Creating 3D models directly from photographs is appealing since it offers the potential of economical acquisition for photorealistic visualization. The landmark method of Debevec et al. [DTM96] produces visually pleasing results for architectural applications. However, detailed surface meso-structure of building materials is rarely considered in such models. To correctly relight different materials requires separation of the way surfaces scatter light and the actual light striking the surface. Although solutions to separate these under specific constraints have been proposed [NVN03], the problem is not generally solvable. To fill in the missing information, humans use tacit knowledge gained from experience of real world illumination to estimate material properties [FDA03]. A number of meso-structure recovery methods capture normal and texture maps with multiple sources [RB99, LKG$^+$03]. An accurate but data intensive approach is to capture and encode the appearance of textured surfaces with a gantry under a large number of lighting and viewing conditions [DvGNK99]. Other methods to recover albedo and meso-structure exist, but require sets of images and/or specialized equipment [YDMH99, LYS01, LFTW06, ND06, PCF05, PC06].

Classic shape-from-shading solutions aim to acquire 3D depth information from a single image [KvD83, Hor89, MM89, HF98a, HF98b, PF05]. This is an under-constrained problem. Numerous shapes, surface reflectances, and lighting conditions

can give rise to the same shading pattern [BKY99], and associated ambiguities in shape perception [Ram88, LB01]. However, shape-from-shading approaches are attractive for our application as they do not require special equipment or lengthy data-capture processes. Khan et al. [KRFB06] successfully demonstrated how, under certain circumstances, limitations in our ability to correctly interpret depth and lighting [OCS05] can be exploited to create plausible synthetic images from a dark-is-deep approximation [LB00]. Our depth hallucination approach is inspired by their ideas.

A large body of literature on the topic of shape-from-shading exists, and we refer to published surveys for a review of existing methods [ZTCS99, DFS07]. Broadly, our approach performs irradiance estimation and is similar in spirit to the iterative technique of Langer and Zucker [LZ94]. Langer and Zucker's model is specifically designed for recovering shape-from-shading on a cloudy day. They observe that under diffuse lighting, surface luminance depends primarily on a local aperture function defined as the solid angle subtended by the visible sky at each surface point. They formulate a set of constraints, applying a robust numerical approach to solve for depth. However, there are a few practical limitations to their method. First, the model assumes uniform albedo, which is a problem for a wide range of textured surfaces. Second, their approach suffers from quantization errors since they perform discretized sampling of light source directions over a hemisphere at each point on a hypothetical surface. Third, their implementation is based on an iterative ray-tracing scheme, which is computationally expensive. Since our goal is to recover sufficient depth for plausibly relighting textured surfaces, we develop a simpler, deterministic image-space solution that approximates their results.

A.4   Depth Hallucination Method

We assume our surface can be plausibly represented as a height field, whose underlying material matrix is approximately Lambertian and opaque, with average reflectance, $2\% < \rho < 70\%$ or so. Our overall process is illustrated in Figure A.2. The individual steps are image capture, albedo and shading estimation, depth estimation, and relighting the surface. Specifics of how we estimate albedo and shading depend on whether the input to our process is a diffuse-lit/flash-lit image pair [ED04], or a single diffuse-lit image. Subtracting the diffuse-lit image from the flash-lit image gives a reasonable estimate of albedo, and a comparison of our diffuse-lit image and albedo provides a

```
┌─────────────────┐        ┌─────────────────┐
│  Flash Image    │        │  Diffuse Image  │
│  (Captured)     │        │  (Captured)     │
└─────────────────┘        └─────────────────┘
        │              ╲           │
        ▼               ╲          ▼
┌─────────────────┐      ╲ ┌──────────────────────┐
│  Albedo Image   │───────▶│ Diffuse Shading Image │
│  (Estimated)    │        │  (Estimated)          │
└─────────────────┘        └──────────────────────┘
        │                           │
        │                           ▼
        │                  ┌─────────────────┐
        │                  │   Depth Map     │
        │                  │  (Estimated)    │
        │                  └─────────────────┘
        │                           │
        ▼                           ▼
      ┌──────────────────────────────┐
      │    Final 3D Surface          │
      │    (Rendered)                │
      └──────────────────────────────┘
```
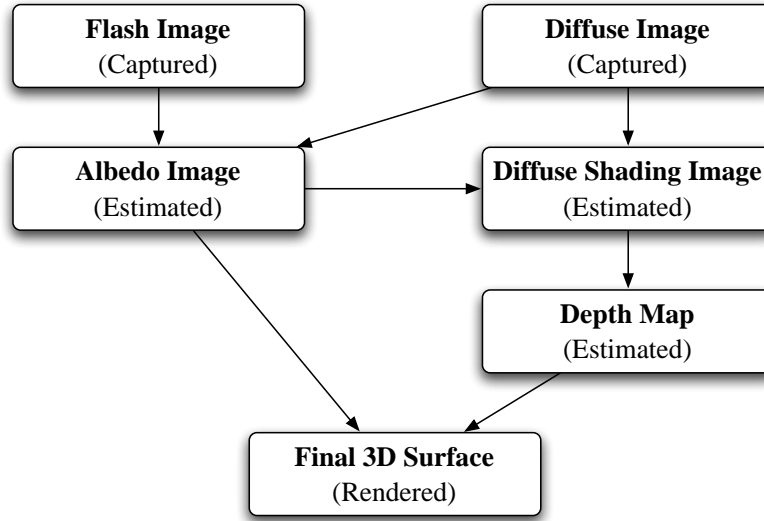
Figure A.2: *Flow chart showing the steps in our process.*

usable estimate of diffuse shading for depth estimation. We discuss this in further detail in Section A.4.2. Our depth estimation method is described in A.4.3, and rendering of our final images is described in A.4.4. Throughout these sections, we illustrate the steps in our process with a case study of a brick path and show the output of each intermediate step.

A.4.1   Image Capture

To capture our input images, we employ a standard digital SLR camera mounted on a tripod, and an attached strobe. Our method requires that we capture a sample of the textured surface without global curvature, as might be found on a wall or floor. If the textured surface contains significant specularities, cross-polarization (i.e., the polarizer on the flash is perpendicular to the polarizer on the lens) can be used to minimize highlights [Her08].

First we capture a RAW format image[1] under indirect illumination (i.e., overcast skies or shadow). We call this the diffuse-lit condition. A second photo is taken from the same point with the flash fired at full power. The camera is set to its maximum flash synchronization speed, while position, aperture, and focus are fixed to ensure good pixel registration between the diffuse-lit and flash-lit conditions. Ideally, the

---

[1] We use RAW format to simplify calibrating the images to each other, however our technique also works with linearized JPEGs.

flash should be mounted as close to the camera lens as possible in order to minimize shadows, though the images shown in this paper were all taken with a standard flash mount. See Figure A.3 for an example input image pair.

A.4.2   Albedo Map and Shading Image

The first stage in our method requires estimation of albedo and diffuse shading. We begin by calibrating our RAW image captures to one another based on their aperture A (f-stop), ISO I, and shutter speed T_s and convert to linear, floating-point pixel values using the following exposure correction factor C_e:

$$C_e = \frac{A^2}{(T_s I)} \tag{A.1}$$

If absolute values were required, there would be an additional conversion factor, which is unnecessary for relative measurements such as ours.

To calculate albedo $I_a(j)$ we perform the operation expressed below at each pixel j:

$$I_a(j) = \frac{I_f(j) - I_d(j)}{I_c(j)} \tag{A.2}$$

Pixel values in the diffuse-lit image I_d are subtracted from our flash-lit capture I_f, and we divide the result by pixel values in the flash calibration image I_c taken of a white Lambertian surface at a similar distance and aperture. This yields approximate reflectance values at each pixel, simultaneously correcting for vignetting, fall-off, and the global cosine factor. Since the cosine factor also depends on the local surface
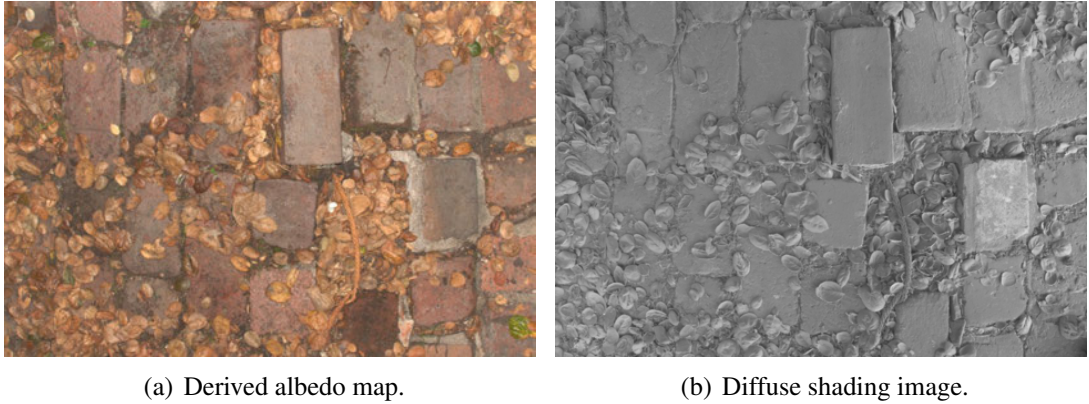

(a) Photograph of a brick path taken in shadowed daylight conditions.

(b) Flash-lit photograph of the brick path.

Figure A.3: *An example input photograph pair.*

(a) Derived albedo map. (b) Diffuse shading image.

Figure A.4: *Example albedo map and shading image generated from the photographs in Figure A.3 of the brick path.*

normal, we may underestimate albedo in steeply sloped areas. In the overall method, such errors will manifest as slight edge shifts, which are very difficult to detect visually. We apply a daylight white balance that provides a good match to the flash, therefore image subtraction results in a good color balance in our albedo image, as shown in Figure A.4(a). In cases where flash shadows are present, we also apply a simple thresholding and neighbor-filling technique that copies detail from the flash-lit areas [PSA+04]. In more severe cases, we can apply an intelligent shadow removal algorithm [FHCD06], though this requires some user intervention. (All examples shown in the paper used the simpler, automatic method.)

To compute the diffuse shading image, we take the ratio of the diffuse-lit condition over the albedo at each pixel. This can result in a color cast due to skylight or cloudy illumination, but our depth estimation method uses only the luminance channel. A computed grayscale shading image for our brick path is shown in Figure A.4. The depth estimation method described in the following section assigns a height of 0 to a pixel intensity of 0.5, so we normalize our shading image to this mean value.

In cases where there are significant differences in meso-structure but similar material properties to a previously captured surface, we can use a diffuse-lit image in conjunction with an existing diffuse-lit/flash-lit pair. We transfer the statistics of the diffuse-lit image to the albedo and the diffuse shading image of the existing exemplar using histogram matching [HB95]. Figure A.5(a) shows an example diffuse-lit capture taken close to the location shown in Figure A.3(a). Figure A.5(b) shows a synthesized diffuse shading image computed by applying histogram matching of Figure A.5(a) to Figure A.4(b). The histogram matching method is especially useful in architectural
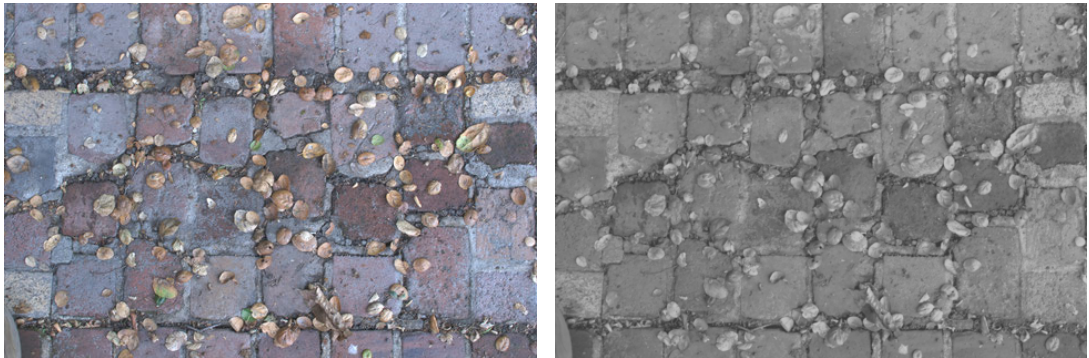
(a) Alternate photograph of a brick path taken in shadow.

(b) Histogram-matched diffuse shading image.

Figure A.5: *Example input image for histogram matching and generated shading image.*

applications, where it is impractical to take flash-lit images of every portion of a large structure, but sample areas with similar appearance and statistics may be readily found.

A.4.3 Depth Estimation

The Langer and Zucker [LZ94] method is designed to recover shape from shading on a cloudy day, which is precisely what we capture in our technique. Applying their relaxation method entails iteratively ray-tracing a discretely sampled hemisphere of light source directions at every surface point. Instead we develop an approximate solution that works entirely in image space and yields a direct estimate of depth at each pixel. A conservative model basis ensures that we do not exaggerate depth variations, and a final, user-specified scale factor achieves the desired roughness.

Surface meso-structure can be approximated as a terrain with hills and valleys. The orientation of the surface to the sky (cosine factor) dominates on the hills, while the visible aperture effect dominates in the valleys, where the sides are at least partly in shadow. We therefore begin by developing two local models to approximate these different types of relationships between meso-structure depth and shading. The scope of each model is shown on a hypothetical textured surface in Figure A.6.

These models are derived such that an above-plane linear model is matched to a below-plane quadratic model at a tangent point, creating the smooth piecewise function plotted in Figure A.7.

Above-plane surface model applies

Below-plane surface model applies
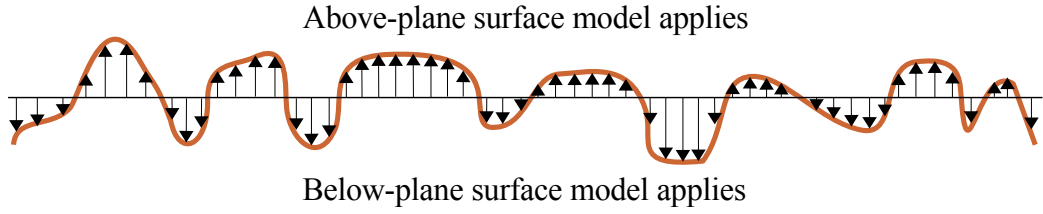
Figure A.6: *Example of a profile of a textured surface and the separation between the above-plane and below-plane surface models.*
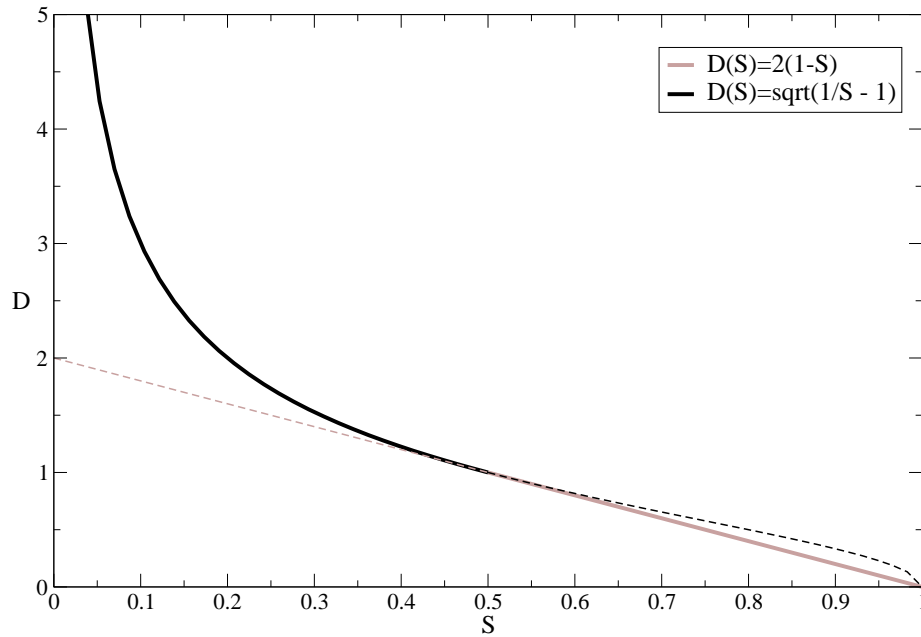
Figure A.7: *The relationship between aperture and shading factor in our model. The dashed line shows the unused extensions of each model.*

Below-Plane Model

We derive our below-plane shadowing model by approximating pits in the surface as cylinders with an aperture $2a$ and depth d, as shown in Figure A.8(a). In order to arrive at a simple formula, we chose to ignore interreflections, which we found affect the scale but not the character of the depth estimates. We calculate an illumination factor E_c by integrating the cosine weighting over the solid angle subtended by the visible sky:

$$E_c = 2\pi \int_0^\theta \cos\theta' \sin\theta' d\theta' = \pi \sin^2\theta \tag{A.3}$$

To arrive at the shading factor S we divide E_c by the illumination factor for the full sky, E_h which can be shown to be π. Through simple trigonometry the integrated shading factor becomes:

$$S = \frac{E_c}{E_h} = \frac{\pi \sin^2 \theta}{\pi} = \frac{a^2}{a^2 + d^2} \tag{A.4}$$

Pit depth can therefore be estimated by solving Equation (A.4) for d as:

$$d = a\sqrt{\frac{1}{S} - 1} \tag{A.5}$$

Above-Plane Model

For the above-plane model, we approximate surface protrusions as hemispheres. Shading of these is a function of the visible portion of the hemisphere h_v subtended by the angle ψ (Figure A.8(b)), and added to the remaining reflected portion of the hemisphere h_r outside this angle.

(a) Cylinder pit model.

(b) Hemisphere protrusion model.
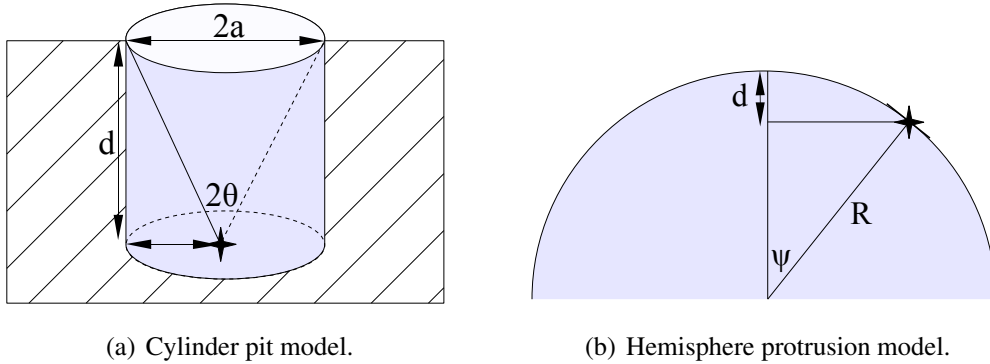
Figure A.8: *Model to approximate shading of pits and surface protrusions.*

Depth can thus be estimated by a simple linear model derived as follows, where ρ is the effective surrounding surface reflectance. (While we will assume $\rho = 0$ in both models, we include it in the above-plane derivation for completeness.)

$$h_v = \frac{\pi}{2}(1 + \cos \psi) \tag{A.6}$$

$$h_r = \rho\frac{\pi}{2}(1 - \cos \psi) \tag{A.7}$$

Consequently our above-plane shading factor is calculated as the ratio of these

quantities and π:

$$S = \frac{\frac{\pi}{2}(1 + \cos\psi) + \rho\frac{\pi}{2}(1 - \cos\psi)}{\pi} \tag{A.8}$$

This can be simplified and solved for $\cos\psi$ to give:

$$\cos\psi = \frac{2S - (1 + \rho)}{1 - \rho} \tag{A.9}$$

From Figure A.8(b):

$$d = R - R\cos\psi \tag{A.10}$$

Substituting $\cos\psi$ gives the linear model:

$$d = 2R\frac{1 - S}{1 - \rho} \tag{A.11}$$

where R is the radius of the hemispherical hill.

Combined Model

These two models, expressed in Equations (A.5) and (A.11) can be conveniently combined at a double root solution to their intersection, by substituting $S = 1/2$ and equating the corresponding values of d:

$$a = \frac{R}{(1 - \rho)} \tag{A.12}$$

Recall, we assume the surrounding surface reflectance $\rho = 0$, yielding depth, d, from the diffuse shading, S, at each scale, a, in the combined aperture formula:

$$D(S) = d/a = \begin{cases} \sqrt{1/S - 1} & \text{for } S \leq 1/2 \\ 2(1 - S) & \text{for } S > 1/2 \end{cases} \tag{A.13}$$

Multiscale Formulation

A shading change over a large region generally corresponds to a greater depth difference than the same shading change over a small region. Since our aperture model estimates depth from diffuse shading relative to a specific feature size, a, we must consider each scale in our captured diffuse shading image separately. Separating our diffuse shading image into scale layers, we efficiently convert our aperture estimates into depth estimates as required for a geometric model.
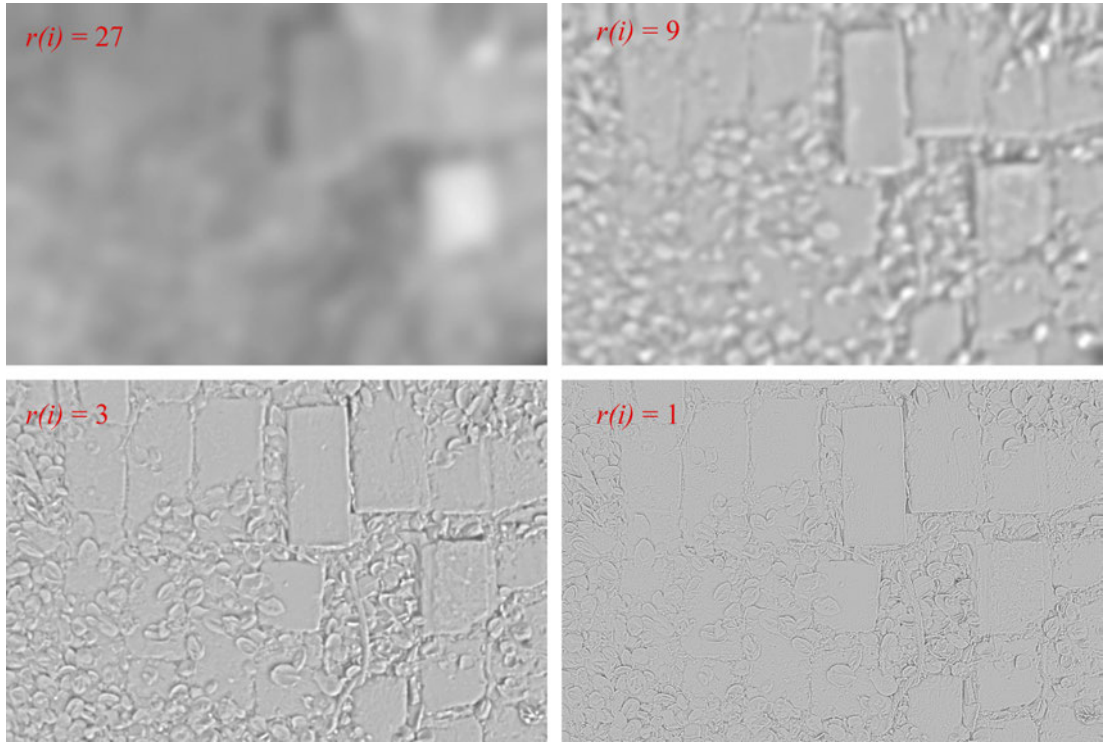
Figure A.9: *The effect of different levels of Gaussian blur on the normalized shading image for the brick path example.*

Starting from a normalized version of our diffuse shading image, shown in Figure A.4(b), we compute several Gaussian blurred images using kernel radii r increasing by powers of three up to a maximum detail size based on image content, which may be specified by the user. At each level, the image is divided by the image at the next largest kernel radius (up to the largest) and multiplied by 1/2 for normalization, effectively yielding a Laplacian pyramid of equal resolution images [BA83a]. These blurred images are referred to as $\ell(i)$. Each of our N levels is then transformed using the depth function D given in Equation (A.13), where a is replaced by the blur radius relative to the synthesized surface size m at each pixel j, arriving at a per pixel depth value D_j:

$$D_j = \sum_{i=1}^{N} \frac{r(i)}{m} \left[D(\ell_j(i)) - 1 \right] \tag{A.14}$$

Figure A.9 shows our progressively blurred shading images for the brick path example. We subtract 1 from our computed depths at each level since this is the normal value for $D(S)$ at the average image intensity of 0.5, and we want our average surface displacement to be zero.

As noted earlier, our depth estimates are conservative. First, we ignored albedo to simplify our analysis. Second, we approximated indentations in the surface as pits, where a crevice model might be more appropriate in some cases. We therefore apply a user-selected, uniform scaling factor to each depth map to compensate for this and achieve an acceptable visual match to the original surface appearance. For all our test scenes, this scaling factor was between 0.75 and 1.5. Our unoptimized implementation takes 15 seconds to generate a 900x600 resolution depth map on a single core 2.5 GHz desktop computer.
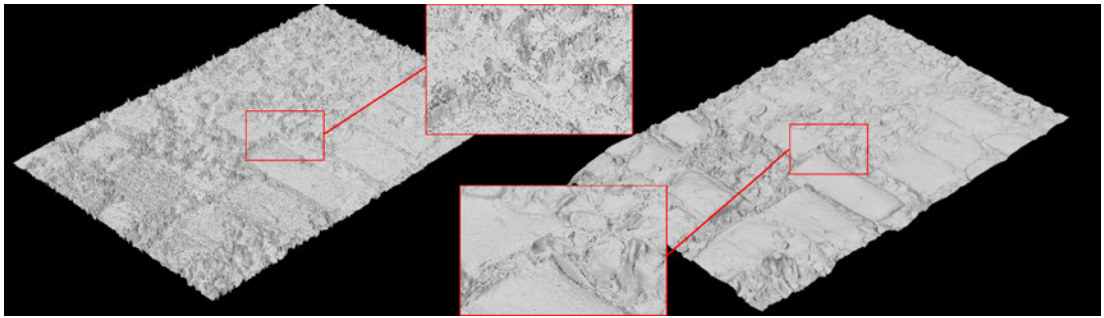
Figure A.10: *A comparison between a simple dark-is-deep approximation (left) and our multiscale model (right).*

In Figure A.10, we show the difference between our depth hallucination method and a global, linear, dark-is-deep approximation [KRFB06] applied to the same diffuse shading image. Notice that our model is less sensitive to noise and better approximates the upper surface as well as the crevices.

(a) Rock wall image.   (b) Depth map obtained using our model.

Figure A.11: *Depth map recovered from a single image of a rock wall obtained from the Web.*

To demonstrate the flexibility of our approach, we downloaded a photograph of a

textured rock wall from an online texture resource[2] taken with an unknown camera, and recovered a depth map using histogram matching to a roughly similar surface to obtain a diffuse shading image. The original image and our hallucinated depth map are shown in Figure A.11.

A.4.4 Relighting the Hallucinated Surface

Once we have an albedo map and a depth map for our surface, virtually any rendering algorithm may be applied. We use the *Radiance* physically-based renderer [War94] with a suitable sky model that includes both direct (solar) and indirect (sky) components, choosing a low angle of solar illumination to make our depth variations more visible. A directly lit surface will have a warmer color cast, and we incorporate this in our model [PSS99]. Figure A.12 shows the results for the brick path example with a solar altitude of $30°$. Specularity is not specifically addressed in our method, but may be added trivially to the material model by assuming a uniform value, as might be encountered on a wet day. (See right-hand image in Figure A.1.) Our validation described in the following section addresses the visual plausibility of our rendered results.

A.5   Experimental Validation

We aim to answer two questions through two experiments. First, can our rendered images be reliably identified as synthetically generated? Second, do renderings generated using hallucinated depth maps appear plausible when compared with renderings using laser-scanned data? If users cannot reliably identify synthetically relit images created using our method while focused on assessing them, then we can conclude our method recovers sufficient detail to allow us to plausibly relight textured surfaces.

Participants with normal vision were seated in front of a standard LCD display. The experimenter ran an application that presented high resolution images to each participant. Depending on the experiment, for each stimulus, participants were asked to press a key to either rank the image or to choose between an image pair. All collected key-presses were logged. We determined the duration for which each image (or pair) was displayed via a pilot study involving 20 participants in which stimuli were presented for 1, 3 and 5 seconds. We found no apparent differences in people's ratings between

---

[2]Source of image: `http://www.texturewarehouse.com`

(a) Original photograph.

(b) Relit brick path.

(c) Histogram-matched relit brick path generated from A.5(a).

(d) Sunny equivalent photograph.
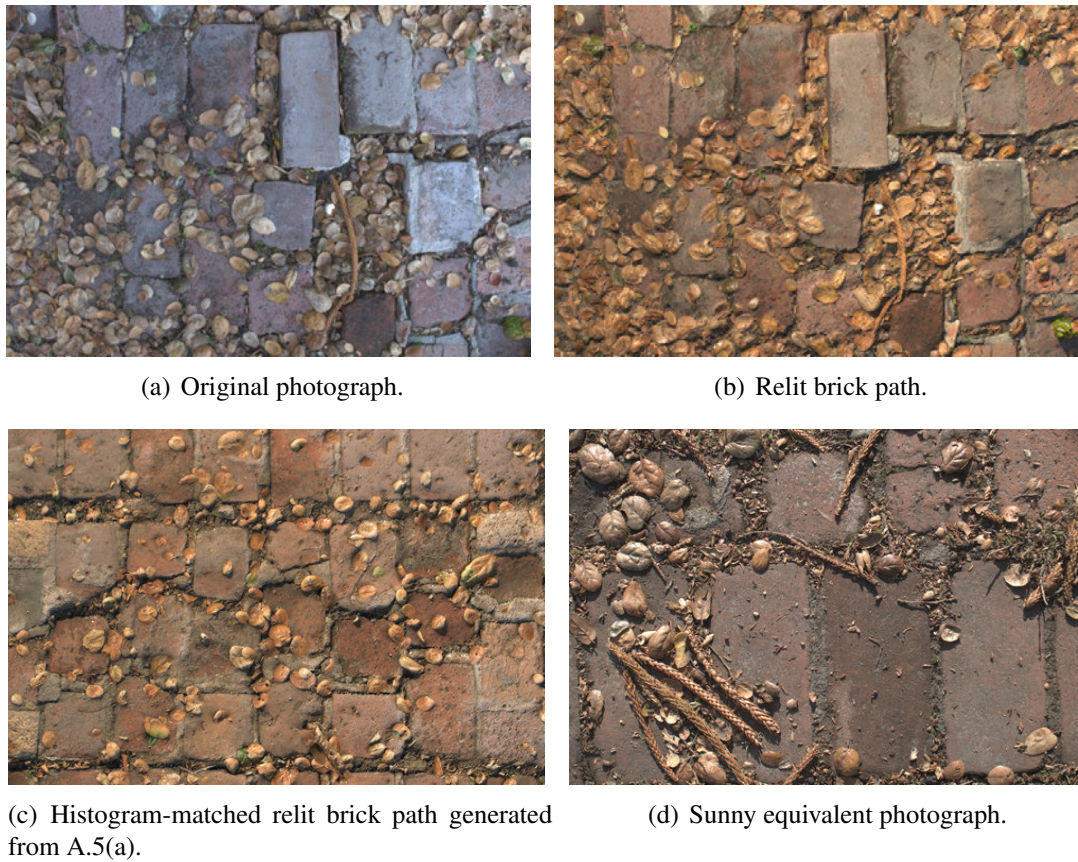
Figure A.12: *Results of relighting our brick path examples.*

(a) Rendering of hallucinated depth map.

(b) Rendering of laser-scanned depth map.
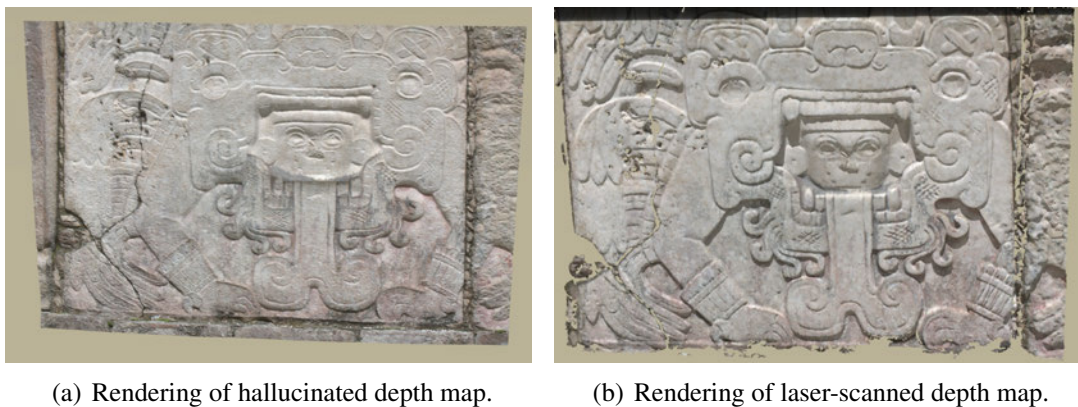
Figure A.13: *Matched lighting frames of the Venus North Platform.*

images shown for given time intervals. Study data was collected from new sets of participants who were shown each stimulus for 3 seconds. A total of 40 participants (20 in each) took part in two studies.

A.5.1 Experiment One

The goal of experiment one was to assess whether people can reliably identify images created using our depth hallucination approach. Single images depicting a variety of textured surfaces, consisting of both real photographs and synthetically relit images, were presented in a randomized order. (See Figure A.1 and Figure A.12 for examples.) A total of 27 images were presented to each person in this part of the study. This set contained 9 day-lit photographs, 9 synthetically relit images and 9 synthetically relit histogram-matched images. Due to the difficulty in acquiring photographs with natural sunny lighting conditions at exactly the same location, the set of equivalent day-lit photographs were not necessarily taken from an identical view point to the images used to recover texture hallucinations. Participants were asked to rank each image from 1 to 5, corresponding to their certainty that the image they were viewing was an untouched photograph. On this scale, we define 1 as definitely synthetically generated, 5 as definitely an untouched photograph, and 3 as undecided.

A.5.2 Experiment Two

Our second study was a two-alternative forced-choice experiment in which the aim was to evaluate the visual plausibility of our estimated depth maps relative to ground-truth data. Twelve pairs of still image frames from an animation depicting changing solar position over the scene with a fixed viewpoint (but novel to the captured one) were used. Each image pair contained an image frame created using ground truth geometry acquired through a laser scanning process, and an equivalent image frame generated using our technique for estimating the depth map from photographs. The same albedo map was registered to both the laser-scan and hallucinated depth maps, and the same physically-based rendering method was used for relighting both sequences. No in-filling techniques were applied to the laser-scan. The lighting frame was matched for specific image pairs (similar to those shown in Figure A.13), but similarly varied between frames in the image sequences. Within each image pair, the laser-scanned and hallucinated surfaces were presented in a randomized order. Participants were asked to choose the image they believed to be most likely to be the real surface. The image pairs shown contained clear visual differences due to the different depth capture processes, but we aimed to answer if they are equally plausible visually.

A.6 Results and Data Analysis

In the first experiment, where participants rated how *real* the images looked, a repeated measures analysis of variance (ANOVA) showed a slight preference for the photographs ($F_{2,38} = 21.61$, p<0.001). This difference was statistically significant. An important result however is that on a scale of 1 to 5: photos received a mean rating of 3.97, relit images scored 3.22 for models derived with diffuse-lit/flash-lit pairs, and 2.98 for histogram-matched versions. The difference between both classes of synthetic images was not found to be statistically significant. On our rating scale a value of above 3 suggests the image is more likely to be a photograph than synthetic.

Relit images were rarely dismissed as artificial, and equivalent photographs were not always recognized as real. For 4 out of our 9 test scenes, the mean scores in Figure A.14(a) show that our synthetic images were virtually indistinguishable from equivalent photographs. In the remaining scenes, relit images were still not rejected outright. Importantly Figure A.14(b) shows that around 15 out of 20 participants gave our synthetic images average ratings above 3, leading us to conclude that our renderings compare very well with photographs. This is further supported by participants commenting in post study de-briefing, on the difficulty in determining which images were synthetic.

In the forced-choice experiment, a paired-sample t-test showed no significant difference between hallucinated depth and the laser-scan. This leads us to conclude that participants could not tell which of the two looked most plausible to them. Mean scores for each choice were 54% for the hallucinated depth, and 46% for the laser-scan. Nine participants out of 20 showed a preference for the hallucinated depth (see Figure A.15), while 6 show a preference for the renderings based on the laser-scanned data. The remaining 5 seem undecided. The viewpoint for each data set was kept identical across all stimuli and close to the captured view to avoid bias for or against either depth map. Within each image pair, the only variable was the depth map used to generate the image. If the view played a significant role in users' assessments then people's choices would have been highly consistent, however only 2 participants chose a particular depth map in every comparison. Between each image pair, lighting was varied. If lighting played a significant role in biasing the results, we would not have seen overall strong differences in preference between participants.
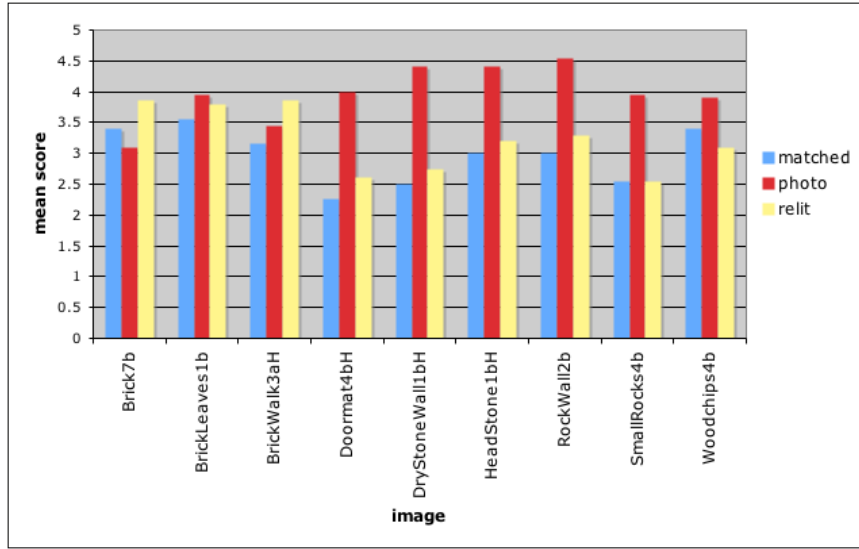
(a) Average ratings from 20 participants, for each specific scene.
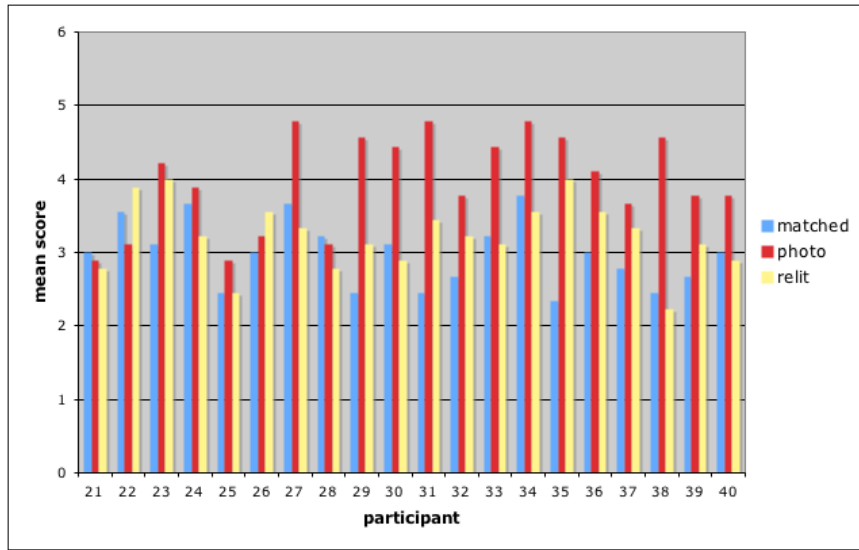
(b) Average ratings per participant, for each class of stimulus.

Figure A.14: *Results from experiment one.*

A.7   Limitations

Naturally, there are situations where our assumptions do not hold, and these may produce unexpected or undesired results. We examine three such cases, which we encountered while acquiring test scenes for our experiments.

The first case is shown in Figure A.16(a), where ivy vines are physically separated from the stone surface below. The separation is small, but it violates one of our basic assumptions, which is that our surface may be plausibly represented as a height field.
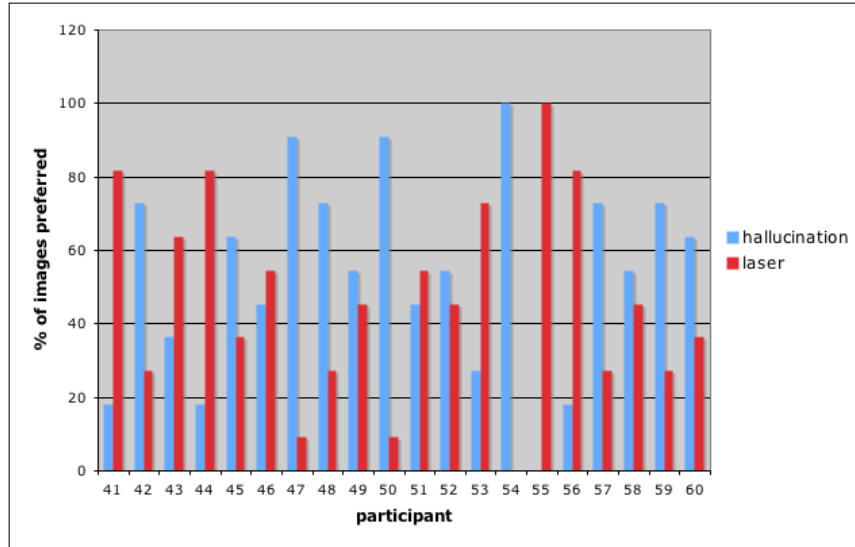
Figure A.15: *Experiment two: percentage of preferred class of stimulus (halluci-nated or laser-scanned) per participant.*

Even if our mathematical model held in this case, which it does not, our height field representation would still fail us. The vines appear to be protruding from the wall rather than next to it, and the rendering looks wrong.

The second case is shown in Figure A.16(b), where our surface is a reasonable match to our geometry assumptions but the daylight illumination is not. In this area, the light comes primarily from one side, as it is nearby a dark structure and only a portion of the sky is visible on the cobbled ground. This results in a bias in the shading image, which our technique translates into a bias in the geometry, making the stones appear to lean towards the original sky direction. While this problem might be overcome with large bounce cards, in a practical setting such biases may be unavoidable and would have to be corrected in a geometry post-processing step.

The third case is shown in Figure A.16(c) and A.16(d). The highly reflective and slightly translucent rock material violated our opaque reflectance assumption, result-ing in a rather flat and unnatural appearance, though some surface structure is still obtained.

A.8   Conclusions

Ultimately, our goal is to combine models of materials approximated using halluci-nated depth maps with wide-baseline 3D reconstructions of buildings. This is likely

(a) Ivy wall relit image.

(b) Cobbles relit image.

(c) Translucent rocks photo.

(d) Relit white rocks.
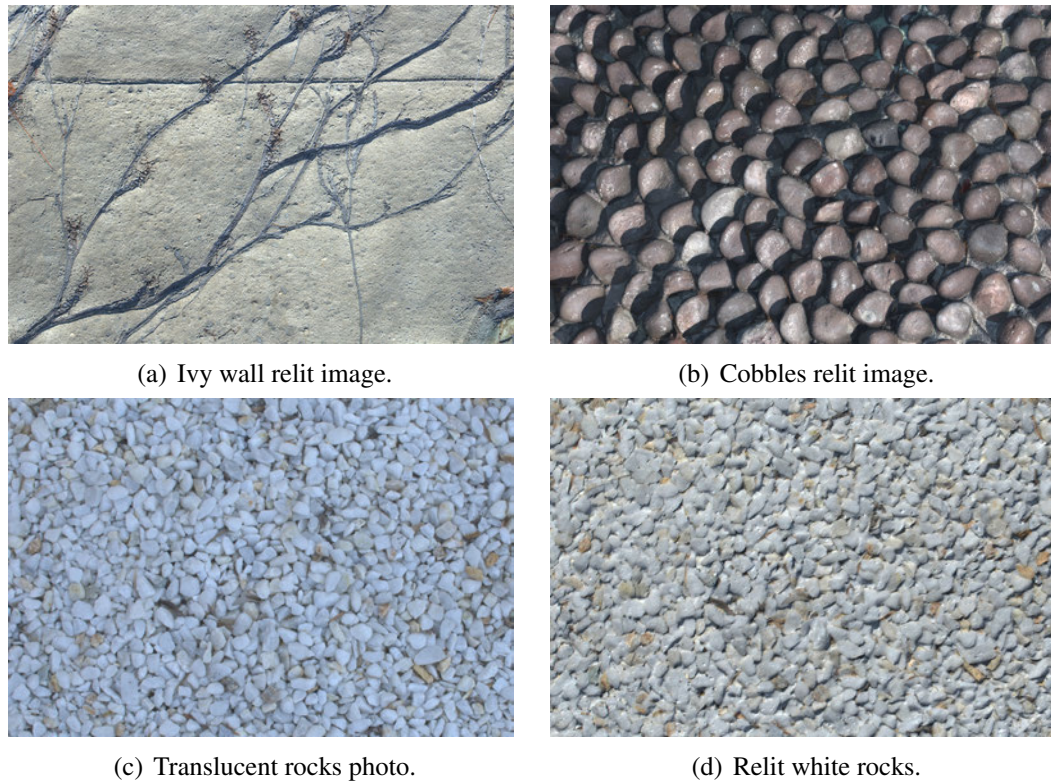
Figure A.16: *Example failure cases that violate fundamental assumptions of our algorithm.*

to pose further challenges, such as surfaces that are unevenly or directly lit, or where neighboring images are taken under differing illumination, or at oblique angles, and must be stitched together seamlessly. Our method complements image-based reconstruction processes by supplying surface detail.

In this work, our objective is to convincingly render the altered appearance of textured surfaces under differing lighting and viewing conditions, while requiring only simple and practical data capture procedures. Starting from diffuse-lit/flash-lit photographs, we generate both an albedo map and textured height field, which can be relit and viewed from any angle under any lighting. Our model applies a surface aperture function but, in contrast to previous methods, works entirely in image space. If only a diffuse-lit image is available, we apply histogram matching with a similar exemplar pair to lift the flash requirement, further simplifying data capture. Compared to alternatives, such as laser scanning, our depth estimation method does not require additional data registration, since both albedo and depth are acquired from perfectly aligned captures.

Histogram matching permits us to hallucinate local height variations from other

diffusely lit imagery, and gaps in our captured model may be filled in using texture synthesis [EF01]. Processing from image to model is also sufficiently simple that depth and albedo maps could be generated on the fly from captured or synthetic texture data on consumer-level graphics cards.

Experimental evaluation of this new approach yielded two important observations. First, when presented with relit images, 75% of participants rated them as more like photographs than synthesized images. Second, participants were unable to decide whether hallucinated depth renderings or those generated using ground truth depth values acquired by laser scanning looked most realistic. Since depth is never fully divulged by shading, our estimates may fall short of absolute accuracy. Our experimental results show that in many practical situations this is unimportant, because the hallucinated depth method reproduces surface appearance that is perceptually tantamount to photographs.

A.9   Acknowledgments

Bibliography

[ADA+04] Aseem Agarwala, Mira Dontcheva, Maneesh Agrawala, Steven Drucker, Alex Colburn, Brian Curless, David Salesin, and Michael Cohen. Interactive digital photomontage. In *SIGGRAPH '04: ACM SIGGRAPH 2004 Papers*, pages 294–302, New York, NY, USA, 2004. ACM. (Cited on pages 113, 114, 115, and 118).

[AF03] A. Agathos and R.B. Fisher. Colour texture fusion of multiple range images. In *3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings. Fourth International Conference on*, pages 139 – 146, 2003. (Cited on page 39).

[Ash01] M. Ashikhmin. Synthesizing natural textures. In *Symposium on Interactive 3D graphics I3D*, pages 217–226. ACM, 2001. (Cited on pages 41, 42, 47, 67, 70, 86, and 90).

[Ash03] Michael Ashikhmin. Fast texture transfer. *IEEE Comput. Graph. Appl.*, 23(4):38–43, 2003. (Cited on pages 41, 47, 86, and 90).

[ASS+09] S. Agarwal, N. Snavely, I. Simon, S.M. Seitz, and R. Szeliski. Building rome in a day. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 72 –79, sep. 2009. (Cited on page 26).

[BA83a] P. Burt and E. Adelson. The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31(4):532–540, 1983. (Cited on page 167).

[BA83b] Peter J. Burt and Edward H. Adelson. A multiresolution spline with application to image mosaics. *ACM Trans. Graph.*, 2:217–236, October 1983. (Cited on page 113).

[Bau02] Adam Baumberg. Blending images for texturing 3d models. In *Proc.*

Conf. on British Machine Vision Association, pages 404–413, 2002. (Cited on pages 111, 112, and 113).

[BG01] S. Boivin and A. Gagalowicz. Image-based rendering of diffuse, specular and glossy surfaces from a single image. In *SIGGRAPH*, pages 107–116, Los Angeles, California, 2001. ACM. (Cited on page 37).

[BIT04] P. Bhat, S. Ingram, and G. Turk. Geometric texture synthesis by example. In *Eurographics/ACM SIGGRAPH symposium on Geometry Processing*, pages 41–44, Nice, France, 2004. ACM. (Cited on pages 42 and 47).

[BJK07] Ronen Basri, David Jacobs, and Ira Kemelmacher. Photometric stereo with general, unknown lighting. *International Journal of Computer Vision*, 72(3):239–257, 2007. (Cited on page 34).

[BKBF02] Alexander Bornik, Konrad Karner, Joachim Bauer, and Franz. High quality texture reconstruction from multiple views. *Journal of Visualisation and Computer Animation*, 12:263–276, 2002. (Cited on pages 114 and 117).

[BKY99] Peter N. Belhumeur, David J. Kriegman, and Alan L. Yuille. The bas-relief ambiguity. *International Journal of Computer Vision*, 35(1):33–44, 1999. (Cited on pages 35 and 159).

[BL04] Kwang-ho Bae and D. D. Lichti. Automated registration of unorganised point clouds from terrestrial laser scanners. In *In: International Archives of Photogrammetry and Remote Sensing, Vol. XXXV, Part B5, Proceedings of the ISPRS working group V/2*, pages 222–227, 2004. (Cited on page 28).

[BL05] M Brown and D. G. Lowe. Unsupervised 3D object recognition and reconstruction in unordered datasets. In *3-D Digital Imaging and Modeling 3DIM*, pages 56–63, 2005. (Cited on pages 26 and 97).

[BM92] Paul J. Besl and Neil D. McKay. A method for registration of 3-d shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2):239–256, 1992. (Cited on page 28).

[BMR01] Fausto Bernardini, Ioana M. Martin, and Holly Rushmeier. High-quality texture reconstruction from multiple scans. *IEEE Transactions on Visualization and Computer Graphics*, 7(4):318–332, 2001. (Cited on pages 111 and 113).

[BR03] E. Beauchesne and S. Roy. Automatic relighting of overlapping textures of a 3d model. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II – 166–73 vol.2, 2003. (Cited on page 39).

[BR07] Benedict J. Brown and Szymon Rusinkiewicz. Global non-rigid alignment of 3-d scans. In *SIGGRAPH '07: ACM SIGGRAPH 2007 papers*, page 21, New York, NY, USA, 2007. ACM. (Cited on page 30).

[BSFG09] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman. Patchmatch: a randomized correspondence algorithm for structural image editing. In *SIGGRAPH*, pages 24:1–24:11. ACM, 2009. (Cited on pages 42 and 70).

[BT78] H.G. Barrow and J.M. Tenenbaum. *Recovering intrinsic scene characteristics from images*. A. Hanson and E. Riseman, editors, Computer Vision Systems. Academic Press, 1978. (Cited on page 35).

[BVZ01] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(11):1222–1239, 2001. (Cited on pages 81, 82, 113, and 117).

[CM91] Y. Chen and G. Medioni. Object modeling by registration of multiple range images. volume 3, pages 2724 –2729, 1991. (Cited on pages 29 and 30).

[CT81] Robert L. Cook and Kenneth E. Torrance. A reflectance model for computer graphics. In *SIGGRAPH '81: Proceedings of the 8th annual conference on Computer graphics and interactive techniques*, pages 307–316, New York, NY, USA, 1981. ACM Press. (Cited on page 36).

[Deb98] P. E. Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *SIGGRAPH*, pages 189–198, Orlando, Florida, 1998. ACM. (Cited on pages 37 and 38).

[DFS07]   Jean-Denis Durou, Maurizio Falcone, and Manuela Sagona. Numerical methods for shape-from-shading: a new survey with benchmarks. *Computer Vision and Image Understanding*, 109(1):22–43, 2007. (Cited on pages 35 and 159).

[DMP+00]   Kristin Dana, Stephen Marschner, Simon Premoze, Holly Rushmeier, Yoichi Sato, and Yizhou Yu. Image-based surface details. ACM SIG-GRAPH Course notes, 2000. (Cited on page 37).

[DSH00]   Yves Dufournaud, Cordelia Schmid, and Radu Horaud. Matching images with different resolutions. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1:612 – 618, 2000. (Cited on page 97).

[DTC04]   A. R. Dick, P. H. S. Torr, and R. Cipolla. Modelling and interpretation of architecture from several images. *International Journal of Computer Vision*, 60(2):111–134, 2004. (Cited on pages 11, 26, 137, and 138).

[DTG+04a]   P. Debevec, C. Tchou, A. Gardner, T. Hawkins, C. Poullis, J. Stumpfel, A. Jones, N. Yun, P. Einarsson, T. Lundgren, M. Fajardo, and P. Martinez. Estimating surface reflectance properties of a complex scene under captured natural illumination. Technical Report ICT-TR-06.2004, University of Southern California Institute for Creative Technologies, 2004. (Cited on page 38).

[DTG+04b]   P. Debevec, C. Tchou, A. Gardner, T. Hawkins, A. Wenger, J. Stumpfel, A. Jones, C. Poullis, N. Yun, P., Einarsson, T. Lundgren, P. Martinez, and M. Fajardo. Estimating surface reflectance properties of a complex scene under captured natural illumination. Technical report, 2004. (Cited on pages 26, 37, and 39).

[DTM96]   Paul E. Debevec, Camillo J. Taylor, and Jitendra Malik. Modeling and rendering architecture from photographs: a hybrid geometry and image-based approach. In *SIGGRAPH*, pages 11–20. ACM, 1996. (Cited on pages 26, 38, and 158).

[DvGNK99]   Kristin J. Dana, Bram van Ginneken, Shree K. Nayar, and Jan J. Koenderink. Reflectance and texture of real-world surfaces. *ACM Transactions on Graphics (TOG)*, 18(1):1–34, 1999. (Cited on page 158).

[DYB98] Paul Debevec, Yizhou Yu, and George Boshokov. Efficient view-dependent image-based rendering with projective texture-mapping. Technical report, University of California at Berkeley, Berkeley, CA, USA, 1998. (Cited on pages 111 and 113).

[ED04] Elmar Eisemann and Frédo Durand. Flash photography enhancement via intrinsic relighting. In *SIGGRAPH*, pages 673–678. ACM, 2004. (Cited on pages 49 and 159).

[EDD⁺95] Matthias Eck, Tony DeRose, Tom Duchamp, Hugues Hoppe, Michael Lounsbery, and Werner Stuetzle. Multiresolution analysis of arbitrary meshes. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, SIGGRAPH '95, pages 173–182, New York, NY, USA, 1995. ACM. (Cited on page 108).

[EF01] Alexei A. Efros and William T. Freeman. Image quilting for texture synthesis and transfer. In *SIGGRAPH*, pages 341–346. ACM, 2001. (Cited on pages 40, 41, 47, 67, 86, 90, and 176).

[EH02] S. F. El-Hakim. Semi-automatic 3D Reconstruction of Occluded and Unmarked Surfaces from Widely Separated Views. In *Proceedings of the ISPRS Commission V Symposium, Close Range Visualization Techniques*, pages 143–148, September 2002. (Cited on page 26).

[EL99] Alexei A. Efros and Thomas K. Leung. Texture synthesis by non-parametric sampling. In *ICCV '99: Proceedings of the International Conference on Computer Vision-Volume 2*, page 1033, Washington, DC, USA, 1999. IEEE Computer Society. (Cited on page 41).

[ELS08] C. Eisenacher, S. Lefebvre, and M. Stamminger. Texture synthesis from photographs. In *Eurographics Conference*, volume 27(2), pages 419–428. Elsevier, 2008. (Cited on pages 121 and 124).

[FB81] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, 1981. (Cited on page 30).

[FCSS10] Y. Furukawa, B. Curless, S.M. Seitz, and R. Szeliski. Towards internet-scale multi-view stereo. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1434 –1441, jun. 2010. (Cited on page 26).

[FDA03] Roland W. Fleming, Ron O. Dror, and Edward H. Adelson. Real-world illumination and the perception of surface reflectance properties. *Journal of Vision*, 3(5):347–368, 2003. (Cited on page 158).

[FH05] Michael S. Floater and Kai Hormann. Surface parameterization: a tutorial and survey. In *In Advances in Multiresolution for Geometric Modelling*, pages 157–186. Springer, 2005. (Cited on page 108).

[FHCD06] G. D. Finlayson, S.D. Hordley, Lu Cheng, and M.S. Drew. On the removal of shadows from images. *Transactions on Pattern Analysis and Machine Intelligence*, 28(1):59–68, 2006. (Cited on page 162).

[Fit03] Andrew W. Fitzgibbon. Robust registration of 2d and 3d point sets. *Image and Vision Computing*, 21(13-14):1145 – 1153, 2003. (Cited on page 28).

[FZ98] Andrew Fitzgibbon and Andrew Zisserman. Automatic 3D model acquisition and generation of new images from video sequences. In *Proceedings of the European Signal Processing Conference*, pages 1261–1269, 1998. (Cited on page 26).

[Geo03] A. S. Georghiades. Recovering 3-D shape and reflectance from a small number of photographs. In *Workshop on Rendering*, pages 230–240, Leuven, Belgium, 2003. Eurographics. (Cited on pages 34 and 37).

[GGH02] Xianfeng Gu, Steven J. Gortler, and Hugues Hoppe. Geometry images. In *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 355–361, New York, NY, USA, 2002. ACM. (Cited on page 126).

[GGV+04] Alberto Guarnieri, Alberto Guarnieri, Antonio Vettore, Antonio Vettore, Fabio Remondino, Fabio Remondino, and Of Pozzoveggiani Church. Photogrammetry and ground-based laser scanning: Assessment of metric accuracy of the 3d model of pozzoveggiani church, 2004. (Cited on page 26).

[GHCH03] S. Gibson, R. J. Hubbold, J. Cook, and T. L. J. Howard. Interactive reconstruction of virtual environments from video sequences. *Computers and Graphics*, 27(2):293–301, 2003. (Cited on pages 26 and 111).

[GKSS02] Igor Guskov, Andrei Khodakovsky, Peter Schröder, and Wim Sweldens. Hybrid meshes: multiresolution using regular and irregular refinement. In *Proceedings of the eighteenth annual symposium on Computational geometry*, SCG '02, pages 264–272, New York, NY, USA, 2002. ACM. (Cited on page 108).

[GP09] Francisco González and Gustavo Patow. Continuity mapping for multi-chart textures. *ACM Trans. Graph.*, 28(5):1–8, 2009. (Cited on page 109).

[GRA09] GRAPHITE. In: http://alice.loria.fr/software/graphite, 2009. (Cited on page 109).

[GSC$^+$07] M. Goesele, N. Snavely, B. Curless, H. Hoppe, and S. M. Seitz. Multi-view stereo for community photo collections. In *International Conference on Computer Vision ICCV*, pages 1–8, 2007. (Cited on page 26).

[GWJ$^+$08] Mashhuda Glencross, Gregory J. Ward, Caroline Jay, Jun Liu, Francho Melendez, and Roger Hubbold. A perceptually validated model for surface depth hallucination. *ACM SIGGRAPH*, 27(3):59:1 – 59:8, 2008. (Cited on pages 11, 22, 35, 36, 37, 38, 40, 43, 46, 49, 56, 66, 99, 138, and 156).

[HB95] David J. Heeger and James R. Bergen. Pyramid-based texture analysis/synthesis. In *SIGGRAPH*, pages 229–238. ACM, 1995. (Cited on pages 41, 53, 55, 56, and 162).

[Her08] Wil Hershberger. Taming those annoying highlights: cross-polarization flash macro photography. Online article, 2008. `http://www.naturescapes.net/042004/wh0404.htm`. (Cited on page 160).

[HF98a] John A. Haddon and David A. Forsyth. Shading primitives: Finding folds and shallow grooves. In *ICCV*, pages 236–241, 1998. (Cited on pages 35 and 158).

[HF98b]   John A. Haddon and David A. Forsyth. Shape representations from shading primitives. In *ECCV '98: Proceedings of the 5th European Conference on Computer Vision-Volume II*, pages 415–431, London, UK, 1998. Springer-Verlag. (Cited on pages 35 and 158).

[HGN01]   Efstathios Hadjidemetriou, Michael D. Grossberg, and Shree K. Nayar. Histogram preserving image transformations. *Int. J. Comput. Vision*, 45(1):5–23, 2001. (Cited on page 123).

[HJO+01]   A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin. Image analogies. In *SIGGRAPH*, pages 327–340. ACM, 2001. (Cited on pages 40, 41, 47, 53, 67, 69, 78, and 89).

[Hor89]   Berthold K. P. Horn. Obtaining shape from shading information. In *Series of Artificial Intelligence: Shape from Shading*, pages 123–171. Mit Press, Cambridge, MA, 1989. (Cited on pages 35 and 158).

[HPS08]   Kai Hormann, Konrad Polthier, and Alia Sheffer. Mesh parameterization: theory and practice. In *ACM SIGGRAPH ASIA 2008 courses*, SIGGRAPH Asia '08, pages 47:1–47:87, New York, NY, USA, 2008. ACM. (Cited on page 108).

[HSNC08]   Byung-Woo Hong, S. Soatto, Kangyu Ni, and T. Chan. The scale of a texture and its application to segmentation. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1 –8, 2008. (Cited on page 98).

[IBG03]   Ryan M. Ismert, Kavita Bala, and Donald P. Greenberg. Detail synthesis for image-based texturing. In *I3D '03: Proceedings of the 2003 symposium on Interactive 3D graphics*, pages 171–175, New York, NY, USA, 2003. ACM. (Cited on pages 56 and 111).

[IGL03]   L. Ikemoto, N. Gelfand, and M. Levoy. A hierarchical method for aligning warped meshes. In *3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings. Fourth International Conference on*, pages 434 – 441, 2003. (Cited on page 30).

[IMNI08]   Katsushi Ikeuchi, Daisuke Miyazaki, Ko Nishino, and Katsushi Ikeuchi. Robust simultaneous registration of multiple range images. In *Digitally Archiving Cultural Objects*, pages 71–88. 2008. (Cited on page 29).

[IOT⁺07] Katsushi Ikeuchi, Takeshi Oishi, Jun Takamatsu, Ryusuke Sagawa, Atsushi Nakazawa, Ryo Kurazume, Ko Nishino, Mawo Kamakura, and Yasuhide Okamoto. The great buddha project: Digitally archiving, restoring, and analyzing cultural heritage objects. *Int. J. Comput. Vision*, 75(1):189–208, 2007. (Cited on page 26).

[IS01] Katsushi Ikeuchi and Yoichi Sato, editors. *Modelling from reality*. Kluwer Academic Publishers, Norwell, MA, USA, 2001. (Cited on page 26).

[JF91] Anil K. Jain and Farshid Farrokhnia. Unsupervised texture segmentation using gabor filters. *Pattern Recognition*, 24(12):1167 – 1186, 1991. (Cited on page 87).

[KBH06] Michael Kazhdan, Matthew Bolitho, and Hugues Hoppe. Poisson surface reconstruction. In *SGP '06: Proceedings of the fourth Eurographics symposium on Geometry processing*, pages 61–70, Aire-la-Ville, Switzerland, Switzerland, 2006. Eurographics Association. (Cited on pages 27 and 32).

[KFCO⁺07] Johannes Kopf, Chi-Wing Fu, Daniel Cohen-Or, Oliver Deussen, Dani Lischinski, and Tien-Tsin Wong. Solid texture synthesis from 2d exemplars. *ACM Trans. Graph.*, 26, July 2007. (Cited on page 70).

[KLMV05] Shankar Krishnan, Pei Yean Lee, John B. Moore, and Suresh Venkatasubramanian. Global registration of multiple 3d point sets via optimization-on-a-manifold. In *Proceedings of the third Eurographics symposium on Geometry processing*, Aire-la-Ville, Switzerland, Switzerland, 2005. Eurographics Association. (Cited on page 29).

[KRFB06] Erum Arif Khan, Erik Reinhard, Roland W. Fleming, and Heinrich H. Bülthoff. Image-based material editing. In *SIGGRAPH*, pages 654–663. ACM, 2006. (Cited on pages 35, 159, and 168).

[KvD83] J. J. Koenderink and A. J. van Doorn. Geometrical modes as a general method to treat diffuse interreflections in radiometry. *Journal of the Optical Society of America JOSA*, 73(6):843–850, June 1983. (Cited on pages 35 and 158).

[KW07]   V. Kwatra and L-Y. Wei. Course 15: Example-based texture synthesis. ACM SIGGRAPH Courses Program, 2007. (Cited on page 66).

[KZN08]   Neeraj Kumar, Li Zhang, and Shree Nayar. What is a good nearest neighbors algorithm for finding similar patches in images? In *Proceedings of the 10th European Conference on Computer Vision: Part II*, pages 364–378, Berlin, Heidelberg, 2008. Springer-Verlag. (Cited on page 70).

[LB00]   Michael S Langer and Heinrich H Bülthoff. Depth discrimination from shading under diffuse lighting. *Perception*, 29(6):649–660, 2000. (Cited on pages 35 and 159).

[LB01]   Michael S Langer and Heinrich H Bülthoff. A prior for global convexity in local shape-from-shading. *Perception*, 30(4):403–410, 2001. (Cited on page 159).

[LFTW06] Hongsong Li, Sing-Choong Foo, Kenneth E. Torrance, and Stephen H. Westin. Automated three-axis gonioreflectometer for computer graphics applications. *Optical Engineering*, 45(4):1–11, 2006. (Cited on page 158).

[LH06]   S. Lefebvre and H. Hoppe. Appearance-space texture synthesis. In *SIGGRAPH*, pages 541–548. ACM, 2006. (Cited on pages 69, 109, and 121).

[LI07]   V. Lempitsky and D. Ivanov. Seamless mosaicing of image-based texture maps. In *Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on*, pages 1–6, June 2007. (Cited on pages 111, 113, 114, and 115).

[Liu07]   Jun Liu. *Automatic Scene Reconstruction from Wide-Baseline Images*. PhD thesis, School of Computer Science, University of Manchester, 2007. (Cited on pages 26, 27, 28, and 31).

[LKG+03] Hendrik P. A. Lensch, Jan Kautz, Michael Goesele, Wolfgang Heidrich, and Hans-Peter Seidel. Image-based reconstruction of spatial appearance and geometric detail. *ACM Transactions on Graphics (TOG)*, 22(2):234–257, 2003. (Cited on page 158).

[LM71]   Edwin H. Land and John J. McCann. Lightness and retinex theory. *J. Opt. Soc. Am.*, 61(1):1–11, January 1971. (Cited on page 128).

[Low04] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004. (Cited on page 27).

[LP92] M.M. Leung and A.M. Peterson. Scale and rotation invariant texture classification. In *Signals, Systems and Computers, 1992. 1992 Conference Record of The Twenty-Sixth Asilomar Conference on*, pages 461 –465 vol.1, October 1992. (Cited on page 97).

[LQ05] Maxime Lhuillier and Long Quan. A quasi-dense approach to surface reconstruction from uncalibrated images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:418–433, 2005. (Cited on page 26).

[LSS⁺98] Aaron W. F. Lee, Wim Sweldens, Peter Schröder, Lawrence Cowsar, and David Dobkin. Maps: Multiresolution adaptive parameterization of surfaces, 1998. (Cited on page 108).

[LW94] Eric P. Lafortune and Yves D. Willems. Using the modified phong brdf for physically based rendering. Technical Report CW197, Computer Science Department, Leuven, Belgium, 1994. (Cited on pages 36 and 37).

[LW08] Guillaume Lavoué and Christian Wolf. Markov random fields for improving 3d mesh analysis and segmentation. *Proceedings of the Eurographics 2008 Workshop on 3D Object Retrieval*, 2008. (Cited on page 114).

[LYS01] Xinguo Liu, Yizhou Yu, and Heung-Yeung Shum. Synthesizing bidirectional texture functions for real-world surfaces. In *SIGGRAPH*, pages 97–106. ACM, 2001. (Cited on page 158).

[LZ94] M. S. Langer and S. W. Zucker. Shape-from-shading on a cloudy day. *Journal of the Optical Society of America*, 11(2):467–478, 1994. (Cited on pages 35, 52, 159, and 163).

[MM89] J. Malik and D. Maydan. Recovering three-dimensional shape from a single image of curved objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6), 1989. (Cited on pages 35 and 158).

[MWLT00] S. R. Marschner, S. H. Westin, E. P. F. Lafortune, and K. E. Torrance. Image-based bidirectional reflectance distribution function measurement. *Applied Optics*, 39(16):2592–2600, 2000. (Cited on page 37).

[ND06] Addy Ngan and Frédo Durand. Statistical acquisition of texture appearance. In Tomas Akenine-Möller and Wolfgang Heidrich, editors, *Eurographics Symposium on Rendering*, pages 31–40. The Eurographics Association, 2006. (Cited on page 158).

[Nis04] David Nistér. Untwisting a projective reconstruction. *International Journal of Computer Vision*, 60:165–183, 2004. (Cited on page 27).

[Nis05] David Nistér. Preemptive ransac for live structure and motion estimation. *Machine Vision and Applications*, 16(5):321–329, 2005. (Cited on page 26).

[NRDR05] Diego Nehab, Szymon Rusinkiewicz, James Davis, and Ravi Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. In *SIGGRAPH '05: ACM SIGGRAPH 2005 Papers*, pages 536–543, New York, NY, USA, 2005. ACM. (Cited on page 126).

[NRH+77] F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis. Geometric considerations and nomenclature for reflectance. Monograph 161, National Bureau of Standards (US), October 1977. (Cited on page 36).

[NVN03] S. G. Narasimhan, R. Visvanathan, and S. K. Nayar. A class of photometric invariants: separating material from shape and illumination. In *ICCV*, pages 1387–1394. IEEE, 2003. (Cited on page 158).

[OCS05] Yuri Ostrovsky, Patrick Cavanagh, and Pawan Sinha. Perceiving illumination inconsistencies in scenes. *Perception*, 34(11):1301–1314, 2005. (Cited on pages 35 and 159).

[OSRW97] Eyal Ofek, Erez Shilat, Ari Rappoport, and Michael Werman. Highlight and reflection-independent multiresolution textures from image sequences. *IEEE Computer Graphics and Applications*, 17, 1997. (Cited on page 114).

[PC06] J. Paterson and K. Cain. Efficient field capture of epigraphy via photometric stereo. In M. Ioannides, D. Arnold, F. Niccolucci, and K. Mania, editors, *Virtual Reality Archaeology and Cultural Heritage (VAST)*, pages 159–161, 2006. (Cited on page 158).

[PCF05] James A. Paterson, David Claus, and Andrew W. Fitzgibbon. BRDF and geometry capture from extended inhomogeneous samples using flash photography. In *Computer Graphics Forum*, volume 24, pages 383–391. Eurographics, 2005. (Cited on pages 34, 37, and 158).

[PF05] E. Prados and O. Faugeras. Shape from shading: a well-posed problem? In *Computer Vision and Pattern Recognition (CVPR)*, pages 870–877. IEEE, 2005. (Cited on pages 35 and 158).

[PGB03] Patrick Pérez, Michel Gangnet, and Andrew Blake. Poisson image editing. In *SIGGRAPH '03: ACM SIGGRAPH 2003 Papers*, pages 313–318, New York, NY, USA, 2003. ACM. (Cited on page 118).

[POF98] Pierre Poulin, Mathieu Ouimet, and Marie-Claude Frasson. Interactively modeling with photogrammetry. In *In Eurographics Workshop on Rendering*, pages 93–104. Springer-Verlag, 1998. (Cited on pages 111, 112, and 113).

[PSA$^+$04] Georg Petschnigg, Richard Szeliski, Maneesh Agrawala, Michael Cohen, Hugues Hoppe, and Kentaro Toyama. Digital photography with flash and no-flash image pairs. In *SIGGRAPH*, pages 664–672. ACM, 2004. (Cited on pages 51 and 162).

[PSS99] A. J. Preetham, Peter Shirley, and Brian Smits. A practical analytic model for daylight. In *SIGGRAPH*, pages 91–100. ACM, 1999. (Cited on page 169).

[PSS01] Emil Praun, Wim Sweldens, and Peter Schröder. Consistent mesh parameterizations, 2001. (Cited on page 108).

[PvV$^+$04] Marc Pollefeys, Luc van Gool, Maarten Vergauwen, Frank Verbiest, Kurt Cornelis, Jan Tops, and Reinhard Koch. Visual modeling with a handheld camera. *International Journal of Computer Vision*, 59(3):207–232, 2004. (Cited on page 26).

[RAKRF08]   Alex Rav-Acha, Pushmeet Kohli, Carsten Rother, and Andrew Fitzgibbon. Unwrap mosaics: a new representation for video editing. In *SIGGRAPH '08: ACM SIGGRAPH 2008 papers*, pages 1–11, New York, NY, USA, 2008. ACM. (Cited on pages 111 and 112).

[Ram88]   V. S. Ramachandran. Perception of shape from shading. *Nature*, (331):163–166, 1988. (Cited on pages 58, 59, and 159).

[RB99]   H. Rushmeier and F. Bernardini. Computing consistent normals and colors from photometric data. In *Second Conference on 3-D Imaging and Modeling 3DIM*, pages 99–108, Ottawa, Canada, 1999. IEEE. (Cited on pages 34 and 158).

[RBMT98]   Holly Rushmeier, Fausto Bernardini, Joshua Mittleman, and Gabriel Taubin. Acquiring input for rendering at appropriate levels of detail: Digitizing a pietà. In *Eurographics Rendering Workshop*, 1998. (Cited on page 26).

[RH01]   Ravi Ramamoorthi and Pat Hanrahan. A signal-processing framework for inverse rendering. In *SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 117–128, New York, NY, USA, 2001. ACM Press. (Cited on page 39).

[RHDG10]   Eric Risser, Charles Han, Rozenn Dahyot, and Eitan Grinspun. Synthesizing structured image hybrids. In *ACM SIGGRAPH 2010 papers*, SIGGRAPH '10, pages 85:1–85:6, New York, NY, USA, 2010. ACM. (Cited on pages 70 and 86).

[RKB04]   Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. "grabcut": Interactive foreground extraction using iterated graph cuts. *ACM SIGGRAPH*, 23(3):309–314, 2004. (Cited on pages 83 and 114).

[RL01]   S. Rusinkiewicz and M. Levoy. Efficient variants of the icp algorithm. In *3-D Digital Imaging and Modeling, 2001. Proceedings. Third International Conference on*, pages 145 –152, 2001. (Cited on page 28).

[RP05]   J. Repko and M Pollefeys. 3D models from extended uncalibrated video sequences: Addressing key-frame selection and projective drift. In *3-D Digital Imaging and Modeling 3DIM*, pages 150–157, 2005. (Cited on page 26).

[SBS05] O. Schall, A. Belyaev, and H.-P. Seidel. Robust filtering of noisy scattered point data. In *Point-Based Graphics, 2005. Eurographics/IEEE VGTC Symposium Proceedings*, pages 71 – 144, jun. 2005. (Cited on page 30).

[Sch94] Christophe Schlick. An inexpensive brdf model for physically-based rendering. *Computer Graphics Forum*, 13(3):233–246, 1994. (Cited on page 36).

[SCSI08] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani. Summarizing visual data using bidirectional similarity. In *Computer Vision and Pattern Recognition CVPR*, pages 1–8. IEEE, 2008. (Cited on page 71).

[SPR06] Alla Sheffer, Emil Praun, and Kenneth Rose. Mesh parameterization methods and their applications. *Found. Trends. Comput. Graph. Vis.*, 2:105–171, January 2006. (Cited on page 108).

[SSS06] Noah Snavely, Steven M. Seitz, and Richard Szeliski. Photo tourism: Exploring photo collections in 3D. *ACM SIGGRAPH*, pages 835–846, 2006. (Cited on page 26).

[SSS+08a] Sudipta N. Sinha, Drew Steedly, Richard Szeliski, Maneesh Agrawala, and Marc Pollefeys. Interactive 3D architectural modeling from unordered photo collections. *ACM SIGGRAPH Asia*, 2008. (Cited on page 26).

[SSS08b] Noah Snavely, Steven M. Seitz, and Richard Szeliski. Modeling the world from internet photo collections. *Int. J. Comput. Vision*, 80(2):189–210, 2008. (Cited on page 26).

[STJ+04] Jessi Stumpfel, Chris Tchou, Andrew Jones, Tim Hawkins, Andreas Wenger, and Paul Debevec. Direct hdr capture of the sun and sky. In *AFRIGRAPH '04: Proceedings of the 3rd international conference on Computer graphics, virtual reality, visualisation and interaction in Africa*, pages 145–149, New York, NY, USA, 2004. ACM. (Cited on page 38).

[SZS+08a] Richard Szeliski, Ramin Zabih, Daniel Scharstein, Olga Veksler, Vladimir Kolmogorov, Aseem Agarwala, Marshall Tappen, and Carsten

Rother. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(6):1068–1080, 2008. (Cited on page 83).

[SZS⁺08b] Richard Szeliski, Ramin Zabih, Daniel Scharstein, Olga Veksler, Vladimir Kolmogorov, Aseem Agarwala, Marshall Tappen, and Carsten Rother. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(6):1068–1080, 2008. (Cited on pages 113, 114, and 117).

[TA08] Alejandro Troccoli and Peter Allen. Building illumination coherent 3d models of large-scale outdoor scenes. *Int. J. Comput. Vision*, 78(2-3):261–280, 2008. (Cited on page 39).

[TA09] S. Todorovic and N. Ahuja. Texel-based texture segmentation. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 841 –848, 2009. (Cited on page 87).

[TFA05] Marshall F. Tappen, William T. Freeman, and Edward H. Adelson. Recovering intrinsic images from a single image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27:1459–1472, 2005. (Cited on page 36).

[TLQ06] Ping Tan, Stephen Lin, and Long Quan. Resolution-enhanced photometric stereo. In *Computer Vision Ű ECCV 2006*, volume 3953 of *Lecture Notes in Computer Science*, pages 58–71. 2006. (Cited on page 34).

[War92] Gregory J. Ward. Measuring and modeling anisotropic reflection. In *SIGGRAPH '92: Proceedings of the 19th annual conference on Computer graphics and interactive techniques*, pages 265–272, New York, NY, USA, 1992. ACM Press. (Cited on page 36).

[War94] Gregory J. Ward. The RADIANCE lighting simulation and rendering system. In *SIGGRAPH*, pages 459–72. ACM, 1994. (Cited on page 169).

[WG09] G. J. Ward and M. Glencross. A case study evaluation: Perceptually accurate textured surface models. In *Symposium on Applied Perception in Graphics and Visualization (APGV)*, pages 109–115. ACM, 2009. (Cited on pages 34 and 37).

[WL00a] Li-Yi Wei and Marc Levoy. Fast texture synthesis using tree-structured vector quantization. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 479–488, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co. (Cited on pages 41, 42, and 69).

[WL00b] Li-Yi Wei and Marc Levoy. Fast texture synthesis using tree-structured vector quantization. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '00, pages 479–488, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co. (Cited on page 70).

[WLKT09] L-Y. Wei, S. Lefebvre, V. Kwatra, and G. Turk. State of the art in example-based texture synthesis. Technical Report EG-STAR, Eurographics State of the Art Report, 2009. (Cited on pages 40, 47, and 66).

[Woo80] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering*, 19(1):139–144, 1980. (Cited on page 34).

[WPH+04] T. Weyrich, M. Pauly, S. Heinzle, S. Scandella, and M. Gross. Postprocessing of scanned 3d surface data. In *Symposium On Point-Based Graphics*, pages 85–94, 2004. (Cited on page 30).

[WSI07] Yonatan Wexler, Eli Shechtman, and Michal Irani. Space-time completion of video. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29:463–476, March 2007. (Cited on page 70).

[WY04] Qing Wu and Yizhou Yu. Feature matching and deformation for texture synthesis. In *ACM SIGGRAPH 2004 Papers*, SIGGRAPH '04, pages 364–367, New York, NY, USA, 2004. ACM. (Cited on page 69).

[XFT+06] Jianxiong Xiao, Tian Fang, Ping Tan, Peng Zhao, Eyal Ofek, and Long Quan. 3d point cloud segmentation for image-based modeling, 2006. (Cited on page 82).

[XFT+08] Jianxiong Xiao, Tian Fang, Ping Tan, Peng Zhao, Eyal Ofek, and Long Quan. Image-based façade modeling. *ACM SIGGRAPH Asia*, 2008. (Cited on page 26).

[XGRD06] Chen Xu, Athinodoros Georghiades, Holly Rushmeier, and Julie Dorsey. A system for reconstructing integrated texture maps for large structures. In *3DPVT '06: Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT'06)*, pages 822–829, Washington, DC, USA, 2006. IEEE Computer Society. (Cited on page 38).

[YDMH99] Yizhou Yu, Paul Debevec, Jitendra Malik, and Tim Hawkins. Inverse global illumination: recovering reflectance models of real scenes from photographs. In *SIGGRAPH*, pages 215–224. ACM, 1999. (Cited on page 158).

[YJL$^+$08] Chih-Chang Yu, Fan-Di Jou, Chun-Chieh Lee, Kuo-Chin Fan, and Thomas C. Chuang. Efficient multi-resolution histogram matching for fast image/video retrieval. *Pattern Recogn. Lett.*, 29:1858–1867, October 2008. (Cited on page 55).

[YLHS05] Hitoshi Yamauchi, Hendrik P. A. Lensch, Jörg Haber, and Hans-Peter Seidel. Textures revisited. *The Visual Computer*, 21(4):217–241, May 2005. (Cited on page 113).

[YM98] Y. Yu and J. Malik. Recovering photometric properties of architectural scenes from photographs. In *SIGGRAPH*, pages 207–217. ACM, 1998. (Cited on pages 38 and 39).

[ZC91] Q. Zheng and R. Chellappa. Estimation of illuminant direction, albedo, and shape from shading. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 13(7):680 –702, jul. 1991. (Cited on pages 35 and 52).

[ZFG08] Hui Zhang, Jason E. Fritts, and Sally A. Goldman. Image segmentation evaluation: A survey of unsupervised methods. *Computer Vision and Image Understanding*, 110(2):260 – 280, 2008. (Cited on page 41).

[ZGWX05] Song-Chun Zhu, Cheng-en Guo, Yizhou Wang, and Zijian Xu. What are textons? *International Journal of Computer Vision*, 62:121–143, 2005. (Cited on page 87).

[Zha92] Zhengyou Zhang. Iterative point matching for registration of free-form curves, 1992. (Cited on page 28).

[ZHW⁺06] K Zhou, X. Huang, Xi. Wang, Y. Tong, M. Desbrun, B. Guo, and H-Y. Shum. Mesh quilting for geometric texture synthesis. In *SIGGRAPH*, pages 690–697. ACM, 2006. (Cited on pages 42 and 47).

[ZSTR07] H. Zhou, J. Sun, G. Turk, and J. M. Rehg. Terrain synthesis from digital elevation models. *IEEE Transactions on Visualization and Computer Graphics*, 13(4):834–848, 2007. (Cited on pages 42 and 47).

[ZTCS99] Ruo Zhang, Ping-Sing Tsai, James Edwin Cryer, and Mubarak Shah. Shape from shading: a survey. *IEEE Transactions on Pattern Analysis and MAchine Intelligence*, 21(8):690–706, 1999. (Cited on pages 35 and 159).

[ZWM98] Song Chun Zhu, Yingnian Wu, and David Mumford. Filters, random fields and maximum entropy (frame): Towards a unified theory for texture modeling. *Int. J. Comput. Vision*, 27(2):107–126, 1998. (Cited on page 41).

[ZZV⁺03] J. Zhang, K. Zhou, L. Velho, B. Guo, and H-Y. Shum. Synthesis of progressively-variant textures on arbitrary surfaces. *ACM Transactions on Graphics*, 22(3):295–302, 2003. (Cited on page 69).