3D Reconstruction for Everyday Things

by

Kai Wu

Bachelor of Engineering, University of Posts and Telecommunications, 2014

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Applied Science

in

THE FACULTY OF APPLIED SCIENCE
(Electric and Computer Engineering Department)

The University of British Columbia
(Vancouver)

April 2017

Abstract

This document provides brief instructions for using the ubcdiss class to write a UBC-conformant dissertation in LaTeX. This document is itself written using the ubcdiss class and is intended to serve as an example of writing a dissertation in LaTeX. This document has embedded Unique Resource Locators (URLS) and is intended to be viewed using a computer-based Portable Document Format (PDF) reader.

Note: Abstracts should generally try to avoid using acronyms.

Note: at University of British Columbia (UBC), both the Graduate and Postdoctoral Studies (GPS) Ph.D. defence programme and the Library's online submission system restricts abstracts to 350 words.

Preface

At UBC, a preface may be required. Be sure to check the GPS guidelines as they may have specific content to be included.

Table of Contents

List of Tables

List of Figures

Glossary

This glossary uses the handy `acroynym` package to automatically maintain the glossary. It uses the package's `printonlyused` option to include only those acronyms explicitly referenced in the LaTeX source.

GPS   Graduate and Postdoctoral Studies

PDF   Portable Document Format

URL   Unique Resource Locator, used to describe a means for obtaining some resource on the world wide web

Acknowledgments

Thank those people who helped you.

  Don't forget your parents or loved ones.

  You may wish to acknowledge your funding sources.

Chapter 1

Related Work

We present in this chapter the related work on appearance modelling and 3D reconstruction techniques. In section 1.1, we provide a comprehensive review of the field of image-based 3D reconstruction based on varied visual cues. In section 1.2, we investigate the contributing factors to object appearance, which serves to choose the best-suited appearance model and reconstruction technique.

1.1 3D Reconstruction Techniques

Image-based 3D reconstruction attempts to recover the geometry of the object or scene from images under different viewpoints or illuminations. The goal of image-based 3D reconstruction can be described as "given a set of images of an object or a scene, estimate the most likely 3D shape that explains those images, under the assumption of known materials, viewpoints, and lighting conditions". This definition reveals that if those assumptions are invaild, this becomes an ill-posed problem since multiple combinations of geometry, viewpoint and illumination can produce exactly the same images [21]. Traditionally, the most common way of dealing with this ambiguity has been to apply smoothness heuristics and regularization techniques [21] to obtain reconstructions that are as smooth as possible. A drawback of this type of approach is that it typically penalizes discontinuities and sharp edges, features that are very common in real scenes.

The 3D reconstruction techniques are typically categorized as passive and ac-

tive methods depending on whether the controlled illumination is required. Passive methods do not require controlled light and can work with ambient light whereas active methods require some form of temporal or spatial modulation of the illumination. We can approach the categorization based on the image cues used to reconstruct the geometry: stereo, shading, contours, texture, defocus, etc. We present different techniques based on the cues exploited in this review.

Three-dimensional model acquisition has always been one of the fundamental research topics in computer vision. Active 3D scanners are currently the dominant technology for capturing digital object models for applications. Their geometric accuracy has continually improved. But they remain expensive, and, more importantly, they suffer from a number of technical limitations. They are invasive and some materials such as hair can not be scanned. They are also not "scalable" to objects of different sizes, especially large ones and outdoor scenes. In comparison, passive image-based modeling from collections of images captured by handheld cameras offers several advantages. It needs only low-cost hardware, it can be applied to objects of any size, and also it preserves the appearance information from original photographs while maintaining perfect geometric alignment.

1.1.1 Stereo Cue

Stereo is one of the most widely used visual cues, and is used in stereoscopy. Stereoscopy estimates the point of a 3D point by triangulation: 1). the corresponding 2D image points are detected and matched across difference views, 2). 3D line containing the center of projection and 2D projection is obtained with known camera parameters, 3). the intersection of all 3D lines is used to recover the 3D point. Trinocular and Multi-View Stereo have been introduced to improve the accuracy and robustness. However these passive approaches suffer from uniform or periodic surfaces. Active approaches overcome this problem using controlled illumination. The active techniques attempt to overcome the correspondence problem by replacing one of the cameras with a controllable illumination source, e.g., single-point laser, slit laser scanner, and temporal or spatially modulated Structured Light (SL), we refer the readers to the survey article by Blais. We discuss various MVS and SL techniques in the current literature.

Multi-View Stereo algorithms can be roughly categorized into four classes: volumetric based, surface evolution based, region growing based, and depthmap based methods [25].

Volumetric methods

The first class computes the cost function in a 3D volume, then extracts a surface from this volume. One successful algorithm is voxel colouring, which traverses a discretized 3D space in depth-order to identify voxels that have a unique colouring, constant across all possible interpretations of the scene. Another thread of work formulates the problem in the Markov Random Field (MRF) framework and extracts the optimal surface by Graph-Cut algorithms.

Seitz et al. proposed a voxel coloring technique that traverses a discretized 3D space in depth- order to identify voxels that have a unique coloring, con- stant across all possible interpretations of the scene.

Surface Evolution

The second class works by iteratively evolving a surface to minimize a cost function. The representations include voxels, level set, and surface meshes. *Space Carving* technique works by iteratively remove inconsisteny voxels from the scene. *Level-set* techniques cast the problem as a variational one, and use a set of PDE's as cost functions, which are deformed from an initial set of surfaces towards the objects to be detected. Other approaches represent the scene as surface meshes that moves as a function of internal and external forces. (Read Hernandez's [8])

The N-view reconstruction problem is generally an ill-posed problem, which means there exists an infinite number of photo-consistent scenes. Kutulakos and Seitz introduced the notion of the *photo hull* and the Space Carving algorithm that computes this least-commitment shape [18]. They can avoid to performing regularization, also ensures that the recovered 3D shape can serve as a description for the entire equivalence class of photo-consistent shapes.

Level-set based techniques minimize a set of partial differential equations defined in a volume. Like space carving methods, level-set methods typically start from a initial volume and shrink inward, or outward if the cost function is mini-

mized. Faugeras and Keriven proposed a novel geometric approach based on variational principle, from which a set of PDE's can be deduced. The level set method is used to deform an initial set of surfaces towards the objects to be detected. However, level-set is no long a popular MVS technique, because high quailty models with correct topology can be directly computed from photo-consistency functions without the refinement steps.

Hiep et al. presented a visibility-based method that transforms a dense point cloud into a surface mesh, which is feed into a mesh-based variational refinement that captures small details, smartly handling photo-consistency, regularization and adaptive resolution.

Region Growing

The third class starts with a sparse set of scene points, and propagate these points to spatial neighbours and refine the cost function with respect to position and orientation. Furukawa and Ponce starts from sparse, reliable seed points, and iteratively expand and filter the set of points to obtain a quasi-dense point cloud. PatchMatch Stereo and the variants start with a randomly initialized 3D volume, and make the assumption that one of the initial patch is close to the "true" one. This true patch can be propagated to spatial neighbours and gets refined to get closer to the optimal patch.

Otto and Chau proposed one of the first work on region growing stereo search. The essence of the algorithm is: start with an approximate match between a point in one image and a point in another, use an adaptive least-squares correlation algorithm to produce a more accurate match, and use this to predict approximate matches for points in the neighbourhood of the first match. Since the stereo matching algorithm is applicable for planar surfaces, it doesn't make sense to match every pixel. Therefore, they first defined a regular grid on the left image, and then defined the "neighbourhood" as four nearest cells in the grid.

Lhuillier and Quan presented a robust two-view quasi-dense correspondence algorithm. They first sort the list of point correspondences using the correlation score, which is called seed points. Then at each step of the propagation, they choose the best corresponding pixels from the list of seed points. Lastly, in the

4

immediate spatial neighbourhood of the selected seed point, they look for new matches and add the bests to the list of seed points according to a combination of local constraints such as correlation, gradient disparity, and confidence. Their approach is the so-call best-first strategy, which can drastically limit the possibility of bad matches and avoid bad initialization.

PMVS is one of the first open source MVS algorithm developed by Furukawa and Ponce. The goal of this method is to reconstruct at least an oriented patch at each grid cell. First, a sparse oriented patch cloud is obtained from triangulating corresponding feature points. At the expansion stage, the current patch with the best Normalized Cross Correlation score is selected and propagated to neighbouring empty cells. Lastly, two visibility-based filtering step are performed to remove erroneous patches lying outside or inside of the "true" surface.

PatchMatch Stereo proposed by Bleyer et al. overcomes a traditional bias that pixels within a support window have the same disparity, or fronto-parallel assumption. The method is inspired by PatchMatch, which is a randomized algorithm for finding approximate nearest neighbour matches between image patches [3]. The method starts by randomly assigning an oriented plane to each pixel in two views. Then each pixel goes through three iterations of propagations and refinement. Each pixel is propagated to the left/top or right/bottom pixels, or corresponding pixel in the second view, or a preceding or consecutive frame for stereo videos. This method can achieve sub-pixel accuracy, but is computational heavy and difficult to parallelism.

There has been some effors extending PatchMatch Stereo to multi-view scenario or proposing new propagation scheme to increase the computational efficiency. A massively parallel method using a diffusion-like propagation scheme was proposed by Galliani et al..

- Galliani, Massively Parallel Multiview Stereopsis by Surface Normal Diffusion

- Uh, Efficient Multiview Stereo by Random-Search and Propagation

- Zheng, PatchMatch Based Joint View Selection and Depthmap Estimation

- Shen, Accurate Multiple View 3D Reconstruction Using Patch-Based Stereo for Large-Scale Scenes

Depthmap Merging

The fourth class works on the image space instead of the scene space, computes a per-view depthmap. By treating a depthmap as a 2D array of 3D points, multiple depthmaps can be considered as a merged 3D point cloud.

This method takes a set of images with camera parameters, discretizes the depth range into a finite set of depth values, then select one with maximum photo-consistency score. Uniform depth sampling may suffice for simple and compact objects. However, for complex and large scenes, a proper sampling scheme is crucial to achieve high speed and quality.

Winner-Takes-All Depthmaps This simple depthmap reconstruction algorithm is to evaluate photo-consistency value throughout the depth range, and pick the depth value with the highest photo-consistency score for each pixel independently. This process is call "Winner-Takes-All" strategy, see figure 1.1.
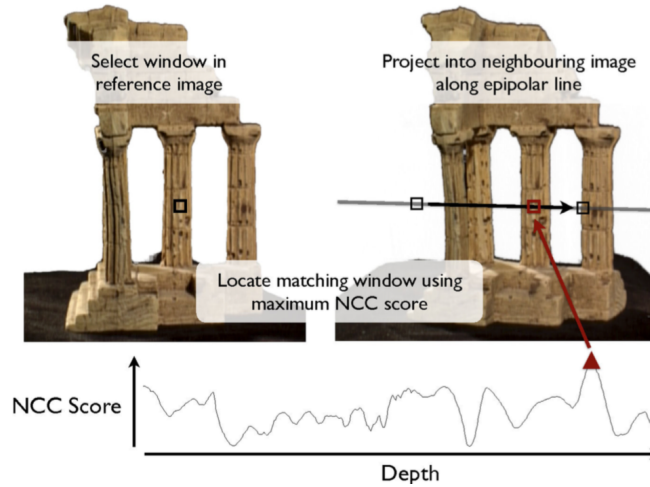

Figure 1.1: Winner-takes-all strategy for depthmap reconstruction.

In addition to the depth value with the highest photo-consistency score, the algorithm often evaluates a confidence measure so that low-confidence depth values can be ignored or down-weighted in the merging step [14]. This algorithm is first proposed in by Esteban and Schmitt.

Though this simplistic approach can in general achieve good enough results, it's still problematic as occlusion or non-Lambertian effects might add noise to the

photo-consistency score. Therefore, a larger window size is more likely leads to a stabler match. However, the associated peak will become broader and less well localized, reducing the accuracy of the depth estimate. Vogiatzis et al. proposed a robust photo-consistency function to overcome this problem. The basic idea is that all potential causes of mismatches like occlusion, image noise, lack of texture, or highlights are treated as outliers. Matching is treated as a problem of robust model fitting to data containing outliers. More explicitly, for each pixel in the reference image, a photo-consistency curve $S_j(d)$ is computed for each visible view $j(j \in \mathcal{N}(i))$. Since simpling avering the photo-consistency scores across various views cannot handle outliers, they build a new $\mathscr{C}$ by detecting all the local maxima d_k of S_j, and using a Parzen window with a kernel W as follows:

$$\mathscr{C}(d) = \sum_{j \in \mathcal{N}(i)} \sum_k S_j(d_k) W(d - d_k) \tag{1.1}$$

This robust photo-consistency score can surpress local mixima, while simple averaging leads to erroneous results, see figure 1.2.

Goesele et al. proposed a simpler yet effective approach, which is to compute the average of pairwise photo-consistency scores after ignoring those below a certain threshold.

MRF Depthmaps In the case of severe occlusion, there may not exist a correspondence in the other images. Spatial consistency can be enforced under the assumption that neighbouring pixels have similar depth values. This can be formulated under the Markov Random Field (MRF) framework, where the problem becomes minimizing the sum of a unary $\Phi(\cdot)$ and pairwise term $\Psi(\cdot, \cdot)$. The unary term is the cost of assigning a depth label k_p from a label set to the pixel p, whereas the pairwise term is the cost of assigning depth label k_p, k_q to a pair of neighbouring pixels p and q, respectively.

$$E(k_p) = \sum_p \Phi(k_p) + \sum_{(p,q) \in \mathcal{N}} \Psi(k_p, k_q) \tag{1.2}$$

The unary cost reflects the photo-consistency score, which in this case, is the inversely proportional to the photo-consistency score. The pairwise term enforces
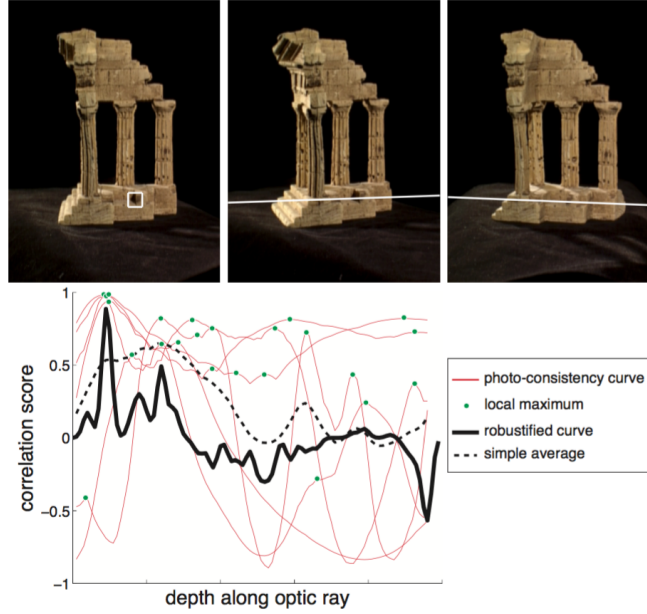
Figure 1.2: Robust photo-consistency function by Vogiatzis et al.

the spatial regularization, thus is proportional to the amount of depth discrepancy at neighbouring pixels.

Structured Light

Structured light is considered one of the most accurate reconstruction technique. It is based on projecting a temporally or spatially modulated pattern onto the surface and viewing the illuminated surface from one or more points of view. The correspondence is easily detected from the projected and imaged pattern, which is triangulated to obtain the 3D point. Each pixel in the pattern is assigned a unique codeword, and the codeword is represented (change of word) by using grey level, colour or geometric representations. Structured light is classified based on the coding strategy: time-multiplex, neighbourhood codification and direct codification [22]. Time-multiplexing techniques generate the codeword by projecting a sequence of patterns. Neighbourhood codification represents the codewords in a unique pattern. Direct codification techniques define a codeword for every pixel, which is equal to its grey level or colour.

Time-multiplexing A sequence of patterns are successively projected onto the surface, the codeword for a given pixel is formed by the sequence of illuminaiton values for that pixel across the projected patterns. This kind of pattern can achieve high accuracy due to two factors: 1). the codeword basis is small, e.g., two for binary pattern, therefore, each bit is easily distinguishable; 2). a coarse-to-fine strategy is used, and the position of the pixel becomes more precise as the patterns are successively projected. We further classify these techniques as follows: 1). binary codeword; 2). *n*-ary codeword; 3). gray code combined with phase shifting; 4). hybrid techniques.

Spatial Neighbourhood This kind of technique concentrate all the coding in a unique pattern. The codeword that labels a certain pixel is obtained from a neighbourhood of the pixels around it. Normally, the visual features gathered in a neighbourhood are the intensity or colour of the pixels or groups of pixels around it.

Direct codification There are ways that can directly represent the codeword in each pixel. To achieve this, there is a need to use either a large range of colour values or introduce periodicity. However, this kind of pattern is highly sensitive to noise because the "distance" between codewords is nearly zero. Moreover, the perceived colour depends not only on the projected colour, but also the intrisic clour of the surface, therefore, reference images must be taken. This kind of coding can be classified as: 1). codification based on grey levels; 2). codification based on colour.

1.1.2 Shading Cue

The shading variations can reveal the surface normal orientation, which can be further integrated into a 2.5D height map. Shading variation depends on the shape (surface normal orientation), reflectance (material), and lighting (illumination), therefore is generally a ill-posed problem because difference shapes illuminated under different light conditions might produce the same image. This leads to a novel technique called Photometric Stereo in which surface orientaiton is determined from two or more images. The idea of Photometric Stereo is to vary the direction of the incident illumination between successive views while holding the viewing direction constant. This provides enough information to determine surface

orientation at each pixel [28]. This technique can produce a surface normal map with the same resolution of the input image, i.e., to produce the pixel-wise surface normal map. Since the coefficients of the normal are continous, the integrated height map can reach an accuracy that cannot be achieve by any triangulation methods. Therefore, photometric stereo is more desirable if the intrisic geometric details are of great importance.

Despite the superior results achieved by Photometric Stereo, traditional photometric stereo generally makes the following assumptions:

- Camera: orthographic projection, linear radiometric response

- Reflectance: known reflectance properties, e.g., Lambertian in [29], specular in [].

- Illumination: the lighting conditions are parallel rays with known directions and intensities.

- Others: shadow, interreflection, and other global light transportation are neglected

The key problem is how to generalize the assumptions of photometric stereo. For the camera assumption, orthographic projection can be achieved by using a lens with long focus and placing the objects far from the camera. The nonlinear response can be solved by performing radiometric calibration. The shadow and other global light transportation are one of the sources of errors, some approaches consider them as outliers and remove them before normal estimation. The reflectance and lighting assumptions, however, are the most complicated ones since the reflectance properties depends on material property and the microscopic structure, and the lighting can have arbitray or fixed position, orientation, and intensity. Therefore the research on Photometric Stereo are generally on three directions: 1). traditional photometric stereo with known reflectance and lighting conditions; 2). generalization of reflectance; 3). generalization of lighting conditions

Traditional case The Photometric Stereo is first proposed in [29]. In this work, the Lambertian model is used, which is constant, and independent to incident and emit light direction.

Generalization of Reflectance refer to [1]

Parametric reflectance model The reflectance is characterized by Bidirectinoal Reflectance Distribution Function (BRDF). Most BRDFs are more complicated than the Lambertian since relative few objects are either ideal diffuse or perfectly specular. The BRDF of many surfaces can be approximated as a combination of a Lambertian component and a specular component. This reflectance model has motivated a line of research [4, 7, 19]. Coleman and Jain and Barsky and Petrou who treat specular pixels as outliers, and Schlüns, Sato and Ikeuchi, and Mallick et al. who assume the color of the specular lobe differs from the color of the diffuse lobe, allowing separation of the specular and diffuse components.

Non-parametric reflectance model While parametric reflectance models are very good at reducing the complexity of BRDFs, they are usually only vaid for a imited class of materials. An alternative is to exploit physical properties common to a large classes of BRDFs. Typical properties include energy conservation, non-negativity, Helmholtz reciprocity, isotropy, etc. Helmholtz stereopsis, introduced by Zickler et al., is on such technique, exploiting reciprocity to obtain surface reconstruction with no dependence to the BRDF. Isotropy is another physical property which holds for material without "grain". Tan et al. use both symmetry and reciprocity present in isotropic BRDFs to resolve the generalized bas-relief ambiguity. Alldrin and Kriegman show that isotropy, with no further assumptions on surface shape or BRDF, can be utilized to recover the surface normal at each surface point up to a plane.

Generalization of Lighting

The generalized lighting condition is anything other than the the ideal case of using a single distant point light source in a dark room. Therefore, any general cases like natural ambient light, multiple point light sources with/without ambient lighting, etc. To make the problem more tractable, the reflectance model should no longer be a general one, otherwise, the problem would have too many degrees of freedom, which means many different shapes with an incorrectly estimated general reflectance, and an incorrectly estimated general lighting would generate the same image appearance with much higher probability.

[Some other papers to read]

- R. J. Woodham. Photometric method for determining sur- face orientation from

multiple images.

1. joint recovery of unknown shape and reflectance

- N. Alldrin, T. Zickler, and D. Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance.

- D. Goldman, B. Curless, A. Hertzmann, and S. Seitz. Shape and spatially-varying BRDFs from photometric stereo.

- T. Higo, Y. Matsushita, and K. Ikeuchi. Consensus photometric stereo.

- B.Shi,P.Tan,Y.Matsushita,and K.Ikeuchi. Elevationangle from reflectance monotonicity: Photometric stereo for general isotropic reflectances.

- BOXIN SHI's PhD thesis

- Satoshi Ikehata's PhD thesis

2. a less restrained capture setup with arbitrary and unknown illumination

- H. Hayakawa. Photometric stereo under a light source with arbitrary motion

- T. Papadhimitri and P. Favaro. A new perspective on uncalibrated photometric stereo

- B. Shi, Y. Matsushita, Y. Wei, C. Xu, and P. Tan. Self-calibrating photometric stereo

3. combine both directions

- A.HertzmannandS.Seitz.Shapeandmaterialsbyexample: a photometric stereo approach.

- A. Hertzmann and S. Seitz. Example-based photometric stereo: shape reconstruction with general, varying BRDFs.

- W. M. Silver. Determining shape and reflectance using mul- tiple images. Masters thesis, MIT, 1980.

1.2   Visual Properties

TBD

Chapter 2

Paper Structure

2.1 Introduction

We want to evaluate the performacne of 4 classes of 3D reconstruction techniques under the condition of different texture, reflectance and shape/geometric properties, including concavity, depth-discontinuity, occlusion.

Contributions

- The first synthetic dataset for 4 different methods comparing their performance under different visual and geometric properties

- An intuitive user interface and high-level abstraction for 3D reconstruction for user without vision background.

2.2 Background

In this section, we discuss the process and assumptions used by the chosen method, the implementation of the chosen algorithms, and justify the choice of the visual and geometric properties.

2.2.1 MVS

We used a seed propagation based implementation. We need to model texture and depth-discontinuity

2.2.2 SL

We projected gray patterns. Need to model Reflectance including albedo (is it caused by albedo, or it's because the color is close to the projected light),

2.2.3 PS

We used Example-based photometric stereo algorithm

2.2.4 VH

We used the image-based visual hull implementation

2.3 Benchmark

In this section, we present a benchmarking of MVS, SL, PS, and VH resulting from our experiments.

2.3.1 Structure of dataset

Shape: plane, cube, sphere

Reflection model: (use disney BSSDF material for rendering), including specular/diffuse ratio (s/d ratio), roughness (from geometric prop), albedo.

Texture model: the most important thing is texture magnitude and orientation, i.e., surface with higher manitude or richer orientation are more likely to be well reconstructed.

Geometric model: roughness, concavity, self-occlusion, depth-discontinuity. Currently, only the first two are considered in the rendering of the synthetic data.

We start by a coarse scale to verify that the performance is consistent among different scales, take reflectance model as an instance:

We've already known that VH works for convex

Properties to be considered:

- texture magnitude

- texture orientation

- albedo

- specular/diffuse ratio

- roughness

- concavity

Solution one:

1). select a specific property, say texture, and use coarse scale(low, medium, and high) for this property, apply (10x10x10) different combinations of (s/d ratio, roughness, albedo). Verify that the performance is consistent among this three cases

2). If it's not consistent, use finer scale property value until a pattern revealed.

We have at least 3(shape)x4(prop num)x3(scale num)x10x10x10 combinations.

Each combination have: 36(MVS+VH)+24(PS)+42(SL)=102 image

Total image number: 36000 * 102 = 3,672,000

Time: 3,672,000 * 1min / 60 / 24 = 61,200 / 24 = 2550 days = 7 years

Storage: $3,672,000 * 100K/10^6 = 367.2GB$

Solution Two:

(1). For each technique, we choose one property at a time while fixing the others, and see if this property will affect the performance of the techniques.

This would result in a cost like:

Combination: 4 * 6 * 10 = 200

Image num: (36 + 24 + 42) * 6 * 10 = 6120

Time: 6120 * 1min / 60 / 24 = 4.25 days

The results might look something like:

MVS

- texture magnitude

- texture orientation

- specular/diffuse ratio

SL

- specular/diffuse ratio

- albedo

- concavity

PS

- specular/diffuse ratio

- albedo

- concavity

VH

- concavity

(2). Not all the properties would have an effect on the specific technique. For each technique, we choose to vary one specific property at a time while fixing the others, use the generated images to evaluate the performance of this technique in terms of this property.

Combination: $10^3 + 10^3 + 10^3 + 10 = 3010$

Images num: $10^3 * (36 + 24 + 42) = 102,000$

Time: $102,000 * 1min/60/24 = 70.8days$

Combination: $5^3 + 5^3 + 5^3 + 5 = 380$

Images num: $5^3 * (36 + 24 + 42) = 12,750$

Time: $12,750 * 1min/60/24 = 8.85days$

We need total of 4.25+8.85 = 13 days just to generate image non-stop.

2.3.2 Experimental setup

2.3.3 Evaluation measures

Accuracy:

    Completion:

2.4 Results

Bibliography

[1] N. Alldrin, T. Zickler, and D. Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008. → pages 11

[2] N. G. Alldrin and D. J. Kriegman. Toward reconstructing surfaces with arbitrary isotropic reflectance: A stratified photometric stereo approach. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE, 2007. → pages 11

[3] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman. PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 28(3), Aug. 2009. → pages 5

[4] S. Barsky and M. Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10): 1239–1252, 2003. → pages 11

[5] F. Blais. Review of 20 years of range sensor development. *Journal of Electronic Imaging*, 13(1), 2004. → pages 2

[6] M. Bleyer, C. Rhemann, and C. Rother. Patchmatch stereo-stereo matching with slanted support windows. In *BMVC*, volume 11, pages 1–11, 2011. → pages 5

[7] E. N. Coleman and R. Jain. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *Computer graphics and image processing*, 18(4):309–328, 1982. → pages 11

[8] C. H. Esteban and F. Schmitt. Silhouette and stereo fusion for 3d object modeling. *Computer Vision and Image Understanding*, 96(3):367–392, 2004. → pages 3, 6

[9] O. Faugeras and R. Keriven. *Variational principles, surface evolution, pde's, level set methods and the stereo problem.* IEEE, 2002. → pages 4

[10] Y. Furukawa and J. Ponce. Accurate, dense, and robust multiview stereopsis. *IEEE transactions on pattern analysis and machine intelligence*, 32(8): 1362–1376, 2010. → pages 4, 5

[11] S. Galliani, K. Lasinger, and K. Schindler. Massively parallel multiview stereopsis by surface normal diffusion. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 873–881, 2015. → pages 5

[12] M. Goesele, B. Curless, and S. M. Seitz. Multi-view stereo revisited. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 2402–2409. IEEE, 2006. → pages 7

[13] V. H. Hiep, R. Keriven, P. Labatut, and J.-P. Pons. Towards high-resolution large-scale multi-view stereo. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1430–1437. IEEE, 2009. → pages 4

[14] X. Hu and P. Mordohai. A quantitative evaluation of confidence measures for stereo vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(11):2121–2133, 2012. → pages 6

[15] K. N. Kutulakos and S. M. Seitz. A theory of shape by space carving. *International Journal of Computer Vision*, 38(3):199–218, 2000. → pages 3

[16] M. Lhuillier and L. Quan. A quasi-dense approach to surface reconstruction from uncalibrated images. *IEEE transactions on pattern analysis and machine intelligence*, 27(3):418–433, 2005. → pages 4

[17] S. P. Mallick, T. E. Zickler, D. J. Kriegman, and P. N. Belhumeur. Beyond lambert: Reconstructing specular surfaces using color. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 619–626. Ieee, 2005. → pages 11

[18] D. Marr. Vision: A computational investigation into the human representation and processing of visual information. 1982. → pages 3

[19] S. K. Nayar, K. Ikeuchi, and T. Kanade. Determining shape and reflectance of hybrid surfaces by photometric sampling. *IEEE Transactions on Robotics and Automation*, 6(4):418–431, 1990. → pages 11

[20] G. P. Otto and T. K. Chau. region-growingalgorithm for matching of terrain images. *Image and vision computing*, 7(2):83–94, 1989. → pages 4

[21] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317(6035):314–319, 1985. → pages 1

[22] J. Salvi, J. Pages, and J. Batlle. Pattern codification strategies in structured light systems. *Pattern recognition*, 37(4):827–849, 2004. → pages 8

[23] Y. Sato and K. Ikeuchi. Temporal-color space analysis of reflection. *JOSA A*, 11(11):2990–3002, 1994. → pages 11

[24] K. Schlüns. Photometric stereo for non-lambertian surfaces using color information. In *International Conference on Computer Analysis of Images and Patterns*, pages 444–451. Springer, 1993. → pages 11

[25] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 519–528. IEEE, 2006. → pages 3

[26] P. Tan, S. P. Mallick, L. Quan, D. J. Kriegman, and T. Zickler. Isotropy, reciprocity and the generalized bas-relief ambiguity. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007. → pages 11

[27] G. Vogiatzis, C. H. Esteban, P. H. Torr, and R. Cipolla. Multiview stereo via volumetric graph-cuts and occlusion robust photo-consistency. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12): 2241–2246, 2007. → pages vii, 7, 8

[28] R. J. Woodham. Photometric stereo: A reflectance map technique for determining surface orientation from image intensity. In *22nd Annual Technical Symposium*, pages 136–143. International Society for Optics and Photonics, 1979. → pages 10

[29] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):191139–191139, 1980. → pages 10

[30] T. E. Zickler, P. N. Belhumeur, and D. J. Kriegman. Helmholtz stereopsis: Exploiting reciprocity for surface reconstruction. *International Journal of Computer Vision*, 49(2-3):215–227, 2002. → pages 11

Appendix A

Supporting Materials

This would be any supporting material not central to the dissertation. For example:

- additional details of methodology and/or data;

- diagrams of specialized equipment developed.;

- copies of questionnaires and survey instruments.