

EUCLIDEAN SHAPE AND POSE RECOVERY OF SURFACES OF REVOLUTION

Cody J. Phillips

A DISSERTATION

in

Computer and Information Science

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2017

Kostas Daniilidis, Professor
Computer and Information Science
Supervisor of Dissertation

Lyle Ungar, Professor
Computer and Information Science
Graduate Group Chairperson

Dissertation Committee

Jean Gallier, Professor
Computer and Information Science
University of Pennsylvania

Camillo J. Taylor, Professor
Computer and Information Science
University of Pennsylvania

Jianbo Shi, Professor
Computer and Information Science
University of Pennsylvania

Kosta Derpanis, Associate Professor
Computer Science
Ryerson University

EUCLIDEAN SHAPE AND POSE RECOVERY OF SURFACES OF REVOLUTION

© COPYRIGHT

2017

Cody J. Phillips

To my mother Marie, from whom I've learned the strength and determination to face life's challenges and opportunities with steadfast resolve and a modicum of grace.

To my father Dave, a man who nurtured my "mechanical" mind and showed me, through his tireless tickering, the power of imagination when combined with the desire of actualization.

*To my husband Tom, who has supported me through the end of this journey, and on through the next.
Oh, the places we will go!*

Acknowledgement

I would like to thank Chuck Stewart and Barb Cutler for their mentorship in my early years of undergrad at RPI. If not for Chuck taking me on as a research student and then encouraging my application to Penn, I might not have found a home at the GRASP Lab under Kostas.

Within my Penn family, I would like to extend my deepest appreciation to my advisor, Kostas Daniilidis, “Papa Kostas”, for all the opportunities he has presented me by way of the many research projects, competitions, and inter-institution collaborations we have tackled through the years. Under his guidance I always found a path forward, as winding as it may be.

For walking this path with me during his years at Penn as a postdoc, I would like to thank Kosta Derpanis. He always goes above-and-beyond in supporting the success of everyone he mentors. His seemingly limitless enthusiasm, optimism, and practical advice was invaluable as I was navigating my first years as a PhD student. Even though his time at Penn has passed, his mentorship continues and he is always there to help critique and improve my work.

As a computer science undergrad, I often viewed the language of formal advanced mathematics as just outside my grasp. A special thanks to Jean Gallier, whose courses in advanced geometric methods provided an immersion (or is it submersion?) in a world of mathematics that had previously given me pause. On days when I found myself lost, just letting the words wash over me like Shakespeare ultimately proved useful in making even the densest of subjects seem more approachable.

I would like to thank my fellow lab mates for all the sanity checks and mini how-to’s over the years. You have never really bonded with someone until you have traced through pages of math trying to find the two sign errors that are cancelling each other out.

Outside of the lab, I would like to thank my husband Tom and my best friend Michael for helping to keep me sane, fed, and appropriately caffeinated during periods of high stress. Their support has been crucial to my progress forward. A final thanks to my dog Sadie, who always keeps me company when I am burning the midnight oil.

ABSTRACT

EUCLIDEAN SHAPE AND POSE RECOVERY OF SURFACES OF REVOLUTION

Cody J. Phillips

Kostas Daniilidis

Surfaces of revolution (SoRs) describe many man-made objects and exhibit several interesting and useful mathematical properties. This thesis explores these relationships from within a Euclidean-based framework and derives minimal problems and algebraic forms for the tasks of single-view and multi-view SoR shape reconstruction, pose recovery, and perceptual grouping. The assumption of a camera with calibrated intrinsic parameters allows projective space to be upgraded to Euclidean space, where image metrology is more readily performed. Specifically, the pose, shape and perspective projection of SoRs are intimately related; knowledge of any two of these three aspects constrains the estimation of the remaining parameters. Four metrology tasks are considered in this dissertation, the first three of which assume known SoR contours. If the SoR shape is known, the absolute pose is recovered from a single view using a one-point minimal correspondence problem (MCP). Both shape and absolute pose are recovered from two extrinsically calibrated views by triangulating the SoR's 3D central axis using estimates of its 2D projection in each view. This two-view triangulation procedure is generalized without the extrinsic calibration in a structure-from-motion (SfM) manner to a two-point MCP to recover the SoR shape and pose — modulo scale. The last metrology task assumes unknown SoR contours. If the SoR pose in n views are known, the SoR geometry permits all views to be mapped into a common shape space. This enables a simultaneous n -view perceptual grouping and shape recovery algorithm. This algorithm is first demonstrated on noise-corrupted SoR views and then applied to a stereoscopic parallax cue that allows the reconstruction of optically challenging SoRs. These methods are validated on real and synthetic datasets.

Contents

Acknowledgement	iv
Abstract	vi
List of Tables	xii
List of Figures	xv
I Introduction and Related Work	1
1 Introduction	2
1.1 Metrology tasks	3
1.1.1 Pose and shape recovery	5
1.1.2 Shape recovery and perceptual grouping	6
1.2 Contributions	7
2 Related work	10
2.1 Apparent contours and contour generators of algebraic surfaces	11
2.2 Straight homogeneous generalized cylinders	11
2.3 Surfaces of revolution under orthographic projection	13
2.4 Surfaces of revolution under perspective projection	14
2.5 Summary	15

II Preliminaries	18
3 Image formation and projection surfaces	19
3.1 Notation and conventions	19
3.2 Image formation	21
3.2.1 Perspective projection and the pinhole camera model	21
3.2.2 Calibrated image plane	21
3.2.3 Coordinate transform	22
3.2.4 Perspective projection	23
3.2.5 Homogeneous image lines and 3D planes	24
3.2.6 Uncalibrated image plane	24
3.3 Projection surface bijections	26
3.3.1 Plane-sphere bijection	27
3.3.2 Plane-plane bijection (two-plane homography)	29
4 Contour generators and surfaces of revolution	31
4.1 General contour generators	32
4.1.1 Apparent contours and projection surfaces	34
4.2 Surface of revolution representation	34
4.2.1 Surface point representation	35
4.2.2 Surface normal representation	37
4.3 Meridian and parallel constraints	38
4.4 Occluding contour and generatrix bijection	39
4.4.1 Canonical camera	40
4.4.2 Center of projection equivalence relation	41
4.4.3 Contour generator from generatrix	42
4.4.4 Contour generator from occluding contour	42
4.4.5 Generatrix from contour generator	44
4.4.6 Generatrix and occluding contour bijection	44

4.5	Forward projection and pose decomposition	45
4.5.1	Representative pose	47
4.5.2	Symmetric pose	47
4.5.3	Canonical pose	47
4.6	Symmetric axis recovery via cylindrical projection	48
III Models and Methods		51
5	Datasets	52
5.1	View-plane dataset	52
5.2	Synthetic view-plane dataset	54
5.3	Transparent stereo dataset	57
6	Symmetry rotation recovery and 3D axis triangulation	59
6.1	Two-point minimal problem for symmetry recovery	59
6.1.1	Two-point minimal problem	60
6.1.2	Minimal problems and parameter estimation	61
6.1.3	Symmetric rotation estimation and refinement	62
6.2	Two-view 3D central axis triangulation	64
6.3	Surface of revolution generatrix recovery	65
6.4	Symmetry rotation recovery evaluation	67
6.4.1	Effect of sensor noise model and iterative refinement	68
6.4.2	Effect of camera position and iterative refinement	68
6.5	Two-view 3D axis triangulation evaluation	71
6.5.1	Effect of sensor noise model	72
6.5.2	Effect of camera position and baseline	72
6.6	Summary	75
7	One-point minimal correspondence problem for absolute pose	76
7.1	One-point correspondence minimal problem	76

7.1.1	Attitude recovery from generatrix derivative	77
7.1.2	Translation recovery from generatrix point	79
7.2	Absolute pose recovery evaluation	80
7.2.1	Ranked hypothesis generation	81
7.2.2	Rank and recall analysis	81
7.2.3	Effect of sensor noise model	83
7.2.4	Effect of camera position	83
7.2.5	Effect of SoR shape	86
7.3	Summary	86
8	Two-point minimal correspondence problem for structure from motion	88
8.1	Two-view system pose parameterization	89
8.2	Two-point minimal correspondence problem	90
8.2.1	Dual attitude recovery	90
8.2.2	Relative translation recovery	92
8.3	Structure-from-motion pose recovery evaluation	93
8.3.1	Effect of sensor noise model	94
8.3.2	Effect of camera position and baseline	94
8.3.3	Effect of SoR shape	99
8.4	Summary	99
9	Simultaneous n-view perceptual grouping and shape recovery	101
9.1	Apparent contour oriented edge response map	102
9.2	Generatrix contour space	104
9.3	Dynamic program optimal subproblem for generatrix optimization	105
9.3.1	Generatrix shape representation	107
9.3.2	Three-dimensional optimal subproblem	108
9.3.3	Four-dimensional optimal subproblem with inflection constraint	108
9.4	Multi-view perceptual grouping and generatrix recovery evaluation	109

9.5	Summary	113
10	Perceptual grouping and shape recovery of optically challenging SoRs	114
10.1	Parallax of a plane	115
10.2	Planar parallax photometric consistency	115
10.3	Parallax cue computation	116
10.4	Non-Lambertian surface properties	118
10.5	Transparent SoR perceptual grouping and generatrix recovery evaluation . .	120
10.6	Summary	124
IV	Discussion and Conclusions	125
10.6.1	Future work	127
V	Appendices	129
	APPENDICES	130
A	Contour generators and surfaces of revolution	130
A.1	Contour generator bijection	130
B	Symmetry rotation recovery and 3D axis triangulation	133
B.1	Symmetry refinement by use of infinitesimal rotations	133
C	Two-point minimal correspondence problem for structure from motion . . .	136
C.1	Dual attitude recovery	136
	Bibliography	139

List of Tables

2.1	Summary of approach assumptions and results	16
3.1	Summary of mathematical terms and conventions	20
5.1	Synthetic contour noise model parameterizations	57
6.1	Effect of iterative refinement on symmetric pose geometric error	68
6.2	Triangulated pose translation and axis estimation errors by dataset	72
7.1	Absolute pose errors by rank and noise model	84
8.1	Two-view SfM pose errors by rank and noise model	95
9.1	Mean SoR reconstruction error by speckle density and image count	111
10.1	Mean SoR reconstruction error by scene and image count	120

List of Figures

1.1	Task one overview, SoR 2D central-axis recovery and 3D triangulation . . .	4
1.2	Task two overview, SoR single-view absolute pose recovery	4
1.3	Task three overview, SoR two-view shape and pose recovery	5
1.4	Task four overview, multi-view SoR perceptual grouping and shape recovery	6
1.5	Task five overview, multi-view generatrix recovery of optically challenging SoRs	7
1.6	Example input images and contours used in prior work	8
2.1	Straight homogeneous generalized cylinders and their subclasses	12
2.2	Special points with projective invariant properties	12
2.3	Effects of orthographic vs. perspective projection	13
3.1	Calibrated camera coordinate system	22
3.2	The back-projection of points and lines	23
3.3	Uncalibrated camera coordinate system	25
3.4	Projective bijections through shared projection center	27
4.1	2D and 3D boundaries of smooth and discontinuous surfaces	32
4.2	General contour generator relationships	33
4.3	Surface of revolution formation	35
4.4	Object and camera coordinate systems	36
4.5	SoR contour generator constraints	38
4.6	Canonical camera pose representation	40
4.7	Projection center equivalence	41
4.8	Forward-projection decomposition	45

4.9	Symmetry of an SoR under cylindrical reprojection	49
5.1	View-plane sample points with respect to the SoR origin	53
5.2	Extremal views from the sampled view plane	53
5.3	Reconstructed SoR generatrices from the view-plane dataset	55
5.4	Synthetic SoR models and corresponding identification numbers	56
5.5	Transparent SoR stereo setup exemplar images	57
6.1	Symmetric points and axes for the symmetry rotation minimal problem . .	60
6.2	Central axis triangulation by intersection of back-projection planes	64
6.3	Metric generatrix reconstruction from occluding contour	66
6.4	Estimated and refined symmetric view-plane axis distance errors	69
6.5	Estimated and refined symmetric view-plane axis-angle errors	70
6.6	Triangulated pose translation and axis estimation errors by camera pose . .	73
6.7	Triangulated pose translation and axis estimation errors by baseline	74
7.1	One-point correspondence for absolute pose	77
7.2	Remapped tangent lines as a function of attitude ϕ	79
7.3	Absolute pose translation and attitude errors by model rank	82
7.4	Absolute pose estimation errors by camera pose	85
7.5	SoR models ranked by mean absolute pose estimate accuracy	86
8.1	SfM two-view system pose parameterization	89
8.2	Two-point correspondence for structure from motion	90
8.3	SfM pose estimation errors by camera pose	96
8.4	Two-view SfM pose estimation errors by camera baseline	97
8.5	SoR models ranked by mean two-view absolute pose estimate accuracy . . .	99
9.1	Three ideal SoR profiles	102
9.2	Three SoR image profiles mapped to generatrix space	102
9.3	Steerable first-derivative Gaussian filter kernels	103

9.4	Steerable first-derivative Gaussian filter basis images and linear combination	104
9.5	Optimal dynamic programming path through the generatrix parameter space	105
9.6	Speckle noise applied to a single view with varying densities	110
9.7	Aggregated SoR responses and reconstructions for varying image counts . .	112
10.1	Geometric consistency and the parallax effect under planar homography . .	116
10.2	Photometric inconsistency of SoR profile under planar homography	116
10.3	Lambertian vs. non-Lambertian surface properties	118
10.4	Planar homography warpings and the parallax cue	119
10.5	SoR parallax responses and reconstructions for a single input image	121
10.6	SoR parallax responses and reconstructions for $k = 10$ input images	122
10.7	SoR parallax responses and reconstructions for varying image counts	123

Part I

Introduction and Related Work

1

Introduction

If I were again beginning my studies, I would follow
the advice of Plato and start with mathematics.

—Galileo Galilei

Perhaps one of the many amazing aspects of human visual cognition is our seemingly innate ability to understand the relationship between an object’s shape, pose, and visual appearance. Looking upon a scene filled with familiar objects, we can intuit each item’s specific shape and placement. In the mind’s eye we can rearrange the world at will. Whether it be a plain black coffee table or a two-tailed purple monkey wearing a fez, our internalized knowledge of shape and pose allows these objects to coalesce into a plausible mental image. What makes this all so remarkable is that the complex mathematical relationships involved in image formation are somehow intricately encoded (or perhaps completely sidestepped) in our visual system. It is this aspect of human perception, the ability to operate over complex mathematical spaces with apparent ease, that presents a persistent challenge to machine perception scientists.

The classical response of the computer vision community to this challenge tends to favor explicit geometric modeling of the imaging process. This is in contrast to the recent trend towards the use of hyperparametric function estimation (e.g., deep learning, end-to-end learning) to abstract away such geometric details. This work follows the former approach. It adheres to the philosophy that explicitly encoding strong geometric constraints at the low

and mid-level vision steps, as appropriate, provides a solid foundation for higher-level visual processing and learning. The models and methods presented in the subsequent chapters demonstrate a degree of elegance that can be achieved by starting with geometric principles.

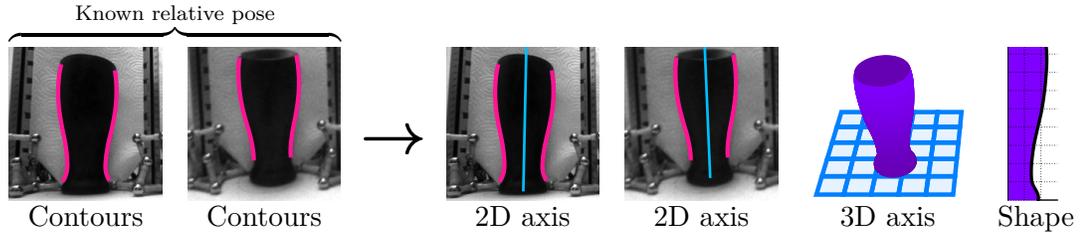
Such principles are developed and laid forth for objects that are defined as surfaces of revolution (SoRs). Like clay vessels formed on a potter’s wheel, surfaces of revolution are created by a contour shape that is swept symmetrically around a central axis. Vases, cups, bowls, and jars – such rotationally symmetric objects have held prevalence in human society since the Early Neolithic period (Derevianko et al., 2004), some 16,000 years ago. They have only received treatment in the computer vision literature however since the early 1990s.

SoRs exhibit interesting mathematical properties and invariants due their inherent rotational symmetry. It is these invariants that are exploited to accomplish various image metrology tasks, such as shape and pose recovery from an SoR image profile. Each research article typically introduces its own SoR geometry formulation with respect to the structural assumptions and invariants (e.g., visible cross sections, bi-tangents points) that underlie its method. The goal of this thesis is to first present a coherent Euclidean framework that relates the most important geometric elements represented throughout the literature. Armed with a comprehensive model of the SoR image formation and reconstruction process, several expansions of prior techniques are accomplished. The aim of the first three presented methods is to distill SoR pose recovery to a series of minimal problems, removing as many structural assumptions as possible. The last two methods provide a means for SoR shape reconstruction on noisy or optically challenging images.

1.1. Metrology tasks

This thesis addresses five primary tasks, the first three (Figures 1.1, 1.2, and 1.3) are related to SoR pose and shape recovery, and the last two (Figures 1.4 and 1.5) are related to SoR shape recovery and perceptual grouping.

Overview:



Input:

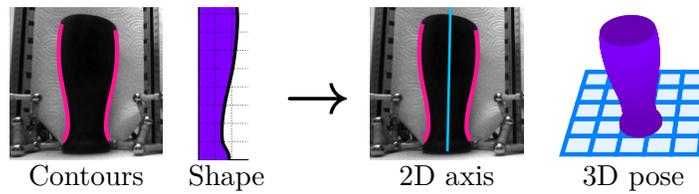
1. Images of an SoR with extracted profile contours from two views
2. Relative transform between the two cameras
3. Camera intrinsic calibrations

Output:

1. SoR 2D central-axis projection in each image
2. SoR 3D central-axis in camera space (modulo height)
3. SoR shape reconstruction (with arbitrary height)

Figure 1.1: Task one overview, SoR 2D central-axis recovery and 3D triangulation.

Overview:



Input:

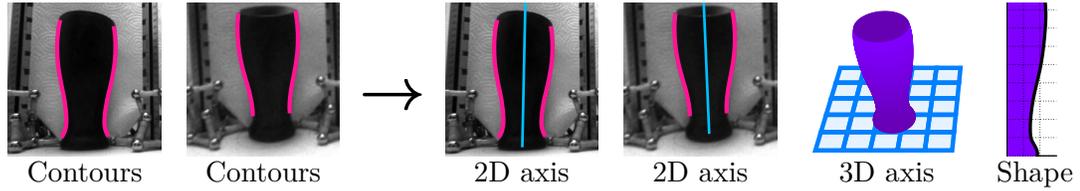
1. Single image of an SoR with extracted profile contours
2. SoR reference shape
3. Camera intrinsic calibration

Output:

1. SoR 2D central-axis projection
2. Absolute pose of the SoR in camera space

Figure 1.2: Task two overview, SoR single-view absolute pose recovery.

Overview:



Input:

1. Images of an SoR with extracted profile contours from two views
2. Camera intrinsic calibration

Output:

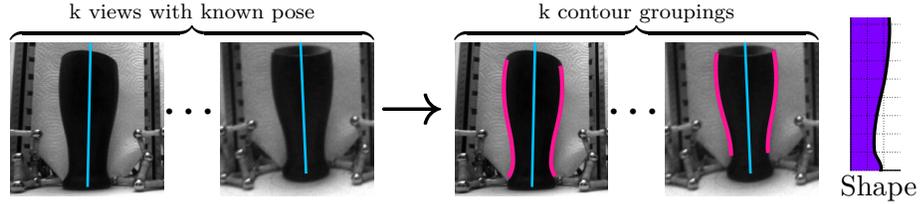
1. SoR 2D central-axis projection in each image
2. Poses of the two SoRs (modulo scale, single-axis motion)
3. SoR shape reconstruction (modulo scale)

Figure 1.3: Task three overview, SoR two-view shape and pose recovery.

1.1.1. Pose and shape recovery

Prior work on SoR pose recovery and shape reconstruction makes several assumptions about the visible SoR image geometry. Removing these constraints is the primary challenge and accomplishment of the first three metrology tasks. The first step in shape reconstruction is pose estimation, which itself first requires the image location of the 2D SoR central-axis projection. Estimating this entity is accomplished in prior work by the use of special projective invariant points. It is computed as part of task one (Figure 1.1) using a two-point minimal problem in a RANSAC-like (Fischler and Bolles, 1981) fashion. Once the 2D central-axis projection is known, its attitude (the amount it tips towards or away from the camera) is needed to recover the 3D central-axis orientation. Prior work requires a visible cross section to estimate this value by parametrizing the projection of the circular cross section to an ellipse. Using two-views with a known relative camera transform, the orientation and position of the 3D central axis is recovered by triangulating the 2D central-axis projections (Figure 1.1). Relaxing the known relative camera transform assumption, the 3D central-axis orientation and position (module-scale) is computed in task three (Figure 1.3) in an SfM fashion using a two-point minimal correspondence problem. In prior work, visible cross

Overview:



Input:

1. *Images of an SoR from multiple camera views*
2. *SoR poses in the camera space*
3. *Camera intrinsic calibration*

Output:

1. *SoR shape reconstruction*
2. *SoR image profile contours*

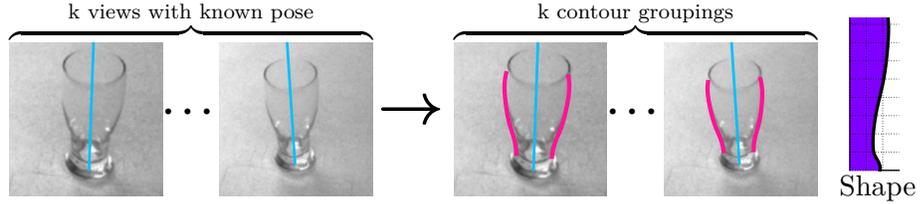
Figure 1.4: Task four overview, multi-view SoR perceptual grouping and shape recovery.

sections and a known shape are used to recover the absolute pose of the SoR with a single view. Problem statement two (Figure 1.2) removes the requirement of visible cross sections and solves for absolute pose using a one-point minimal correspondence problem.

1.1.2. *Shape recovery and perceptual grouping*

Prior work assumes that a high-resolution and readily extractable SoR profile contour is present in the image, as shown in Figure 1.6. This is important as both the pose estimation and reconstruction techniques require well-localized contour edge points with accurate tangent information. No previous work is known to address the case of extreme contour noise and clutter, which has the potential to yield invalid shape reconstructions. This is true even when the SoR pose is known if additional global constraints are not incorporated into the solution. The primary challenge of the last two tasks is to achieve accurate reconstruction results given known SoR poses in the presence of extreme noise using multiple image views. Task four (Figure 1.4) computes the apparent contour grouping and shape reconstruction assuming a known SoR pose, while task five (Figure 1.5) extends this approach to handle transparent and optically challenging SoRs.

Overview:



Input:

1. *Images of an optically challenging SoR from multiple camera views*
2. *SoR poses in the camera space*
3. *SoR supporting plane location*
4. *Camera intrinsic calibration*

Output:

1. *SoR shape reconstruction*
2. *SoR image profile contours*

Figure 1.5: Task five overview, multi-view generatrix recovery of optically challenging SoRs.

1.2. Contributions

This thesis presents an explicitly parameterized Euclidean decomposition of the SoR forward-projection equation and bijection between the shape of an SoR and its image profile. Many of the various components of this formulation are similarly derived or utilized individually in previous work, however they have never appeared together as a coherent framework. It is through this framework that novel geometric relationships and parameterizations are realized and employed to achieve new results.

Such parameterized forms are used to derive minimal problems that overcome the limitations of prior approaches for SoR pose and shape recovery from known occluding contours, in both monocular and stereo settings, as outlined in problem statements 1–3. The SoR shape-profile bijection is used to novelly address the challenges of perceptual grouping and shape recovery for noise-corrupted and non-salient SoR profiles, as summarized in problem statements 3–4.



Figure 1.6: Example input images and contours used in prior work. Prior work generally assumes high-resolution images with properties that enable a reliable SoR profile contour extraction. From left to right, images are excerpted from Dhome et al. (1990), Wong et al. (2002), Utcke and Zisserman (2003), and Colombo et al. (2004).

This thesis allows for full 5-DoF pose and metric generatrix shape recovery (up to a scalar in some instances) without the use of any cross sections or special projective invariant contour points, as employed extensively in prior approaches. This is made possible by the exploitation of the additional geometry constraints revealed in the Euclidean decomposition, and the introduction of a stereo view in some cases.

Contributions by chapter are summarized as follows:

- Chapter 4 presents the pose-parameterized Euclidean SoR forward-projection decomposition along with the SoR profile – shape bijection that in conjunction relate imaged SoR profiles to their pose and shape;
- Chapter 5 presents the datasets collected and used to validate the approaches developed in this thesis;
- Chapter 6 first presents the two-point minimal problem for recovering the image pro-

jection of an SoR's central axis. It then presents how to triangulate the 3D central axis from two such imaged axes of a stereo image pair;

- Chapter 7 presents the one-point minimal correspondence problem for absolute pose recovery, yielding the three pose parameters of camera height, distance, and attitude;
- Chapter 8 presents the two-point minimal correspondence problem for two-view structure from motion, recovering the relative height and depth between two cameras as well as the attitudes of the camera pair;
- Chapter 9 presents the generatrix contour space volume and the dynamic programming optimal subproblem that are used together to accomplish simultaneous n -view perceptual grouping and generatrix recovery;
- Chapter 10 presents a stereoscopic parallax cue that is used in conjunction with the techniques of Chapter 9 to perform perceptual grouping and generatrix recovery on SoRs with optically challenging surface properties.

2

Related work

The SoR forward-projection decomposition and profile–shape bijection formulated in this thesis draw from a body of prior research that began in the late 1980s. The relationship between algebraic surfaces and their image profile under perspective projection is first explored in Giblin and Weiss (1987). This work is specialized to *Straight Homogeneous Generalized Cylinders (SHGCs)*, a geometric superclass of SoRs, in Ponce et al. (1989). The first notable treatment of SoRs is in Dhome et al. (1990) under the assumption of orthographic projection.

In Glachet et al. (1992), the orthographic assumption is relaxed and the earliest perspective SoR forward-projection equation is introduced. Special projective invariant SoR points are described in Forsyth et al. (1992), yielding insight into the projective nature of SoRs. The profile of an SoR is understood to be related by a planar harmonic homology in Zisserman et al. (1995), providing an important mathematical relationship for interpreting SoRs in projective space.

The first equation for metric SoR reconstruction from a projective image is presented in Wong et al. (2002). The authors continue to explore the properties of SoRs under projection, and develop a means of camera calibration from SoR views in Wong et al. (2003). Continuing the trend of projective SoR treatment, another calibration procedure is presented in Colombo et al. (2002), which is expanded upon to allow for a SoR metric reconstruction technique that is defined almost exclusively in projective space. Utcke and Zisserman (2003) presents a SoR reconstruction technique that does not require camera calibration,

but yields a projective family of reconstructions instead of single metric solution.

This chapter traces the history of SoRs and highlights the starting points for this thesis.

2.1. Apparent contours and contour generators of algebraic surfaces

The image profile of a smooth surface corresponds to a viewpoint-dependant 3D curve along such surface. This image profile is called an *apparent contour*, and the corresponding 3D curve is called the *contour generator*. The relationship between these curves and the camera center captures the most general constraints between an object’s shape, pose, and visual appearance for smooth algebraic surfaces. These constraints are developed in Giblin and Weiss (1987) and further expanded in Cipolla (1991) and Cipolla and Blake (1992) in the context of smooth algebraic surface reconstruction for an object observed under motion. These works assume that the camera centers are known in world coordinates, and that the 3D image rays of the apparent contours are also known, i.e., that the camera is calibrated.

There are two results in particular from these works that are pivotal to the development of the SoR forward-projection decomposition and profile-shape bijection. First is the fact that the contour generator and corresponding apparent contours are a sole function of the surface definition and the camera center. Second is the surface tangency and intersection constraint, which states that an apparent contour ray must be contained within the tangent plane of the contour generator point on the surface it intersects.

2.2. Straight homogeneous generalized cylinders

Surfaces of revolutions (SoRs) belong to a central-axis-based class of geometric surfaces known as *straight homogeneous generalized cylinders (SHGCs)*. Such surfaces are defined by an arbitrary cross section \mathbf{C} that is swept along a perpendicular axis $\hat{\mathbf{z}}$, while being scaled by *scaling function* $r(z)$. By this definition, an infinite unit cylinder is the simplest SHGC, consisting of a unit circle cross section and a constant scaling function, $r(z) = 1$. SoRs are a generalization of the cylinder, in that the cross section radius varies arbitrarily

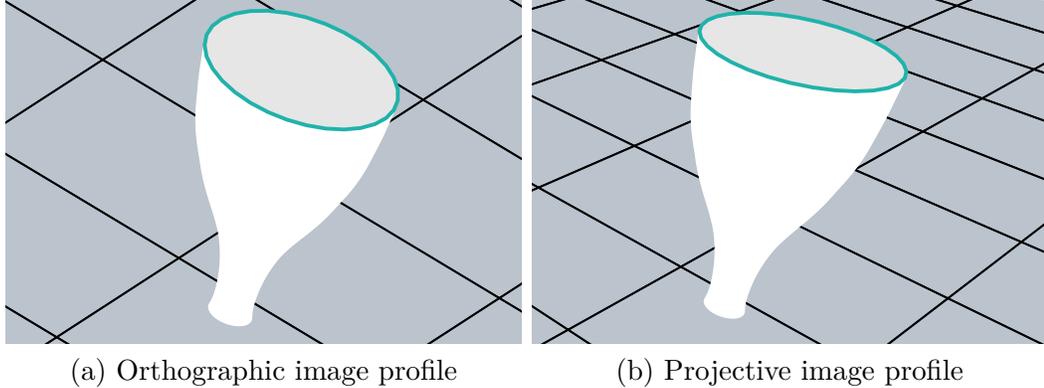


Figure 2.3: Effects of orthographic vs. perspective projection. A near-orthographic camera (a) produces a pseudo-symmetric occluding contour, while a projective camera (b) produces occluding contours that are related by a more general harmonic homology.

When reasoning about the shape of an imaged SHGC, visible cross sections are vital to recovering the unknown cross section shape \mathbf{C} as well as the axis $\hat{\mathbf{z}}$ orientation. For an imaged SoR, cross sections provide even more orientation information as they are known to be projections of a circle. As a testament to this fact, all work prior to this thesis uses one or more visible cross sections to enable pose and shape recovery of SoRs. Gross and Boulton (1996) and Sato and Binford (1992, 1993) present approximate SHGC (and SoR by extension) reconstruction techniques that operate on SHGC profiles created by near-orthographic projection.

2.3. Surfaces of revolution under orthographic projection

The projective treatment of both SoRs and SHGCs was initially limited to orthographic or near-orthographic perspective projections. The relationship between image profile and surface shape is greatly simplified without the effects of perspective distortion. The occluding contour of an SoR under orthographic projection has 2D symmetry about the projection of the central axis $\hat{\mathbf{z}}$ (Figure 2.3a) and corresponds to a coplanar symmetric contour generator.

Dhome et al. (1990) is the first to present an algorithm for SoR pose recovery from image profile contours. The central axis orientation is recovered from a visible cross section. The pose is refined by iterative matching to a triangulated 3D mesh. Glachet et al. (1991)

searches for the 2D central-axis projection in a near-orthographic view by assuming the apparent contours are related by a pseudo-symmetry and verifying that profile points and tangents approximately map onto themselves through an axis reflection.

2.4. Surfaces of revolution under perspective projection

Under perspective projection, the 2D symmetry of the orthographic assumption is replaced with a more general 4-DoF *planar harmonic homology* (Figure 2.3b) and the planar contour generator is replaced with a more irregular surface curve that is dependant upon the surface normal and camera center.

The equation for the SoR contour generator under perspective projection is developed in Glachet et al. (1992). This provides the starting point for the SoR forward-projection equation as well as the shape-to-profile half of the SoR profile-shape bijection. The apparent contours of an SoR can be rendered directly from this equation using its shape representation, a 2D-curve, instead of an approximate 3D mesh model. The SoR pose is computed in this extension to Dhome et al. (1990) the same way as in the original paper, except that the explicit apparent contour equation is used instead of the mesh model. A visible cross-section is used to get the approximate attitude, and a pseudo-symmetry axis search or projective invariant points are recommended as alternative ways of approximating the 2D central-axis projection.

Zisserman et al. (1995) provides a survey of class-based grouping techniques that includes the precise mathematical relationship between two SoR apparent contour pairs under perspective projection. Instead of a symmetry, SoR contours are related by a 4-DoF planar harmonic homology that is often estimated using bi-tangent points. This fact is used heavily in Mendonca et al. (2001); Wong et al. (2002, 2003), where the projective relationships of SoRs under perspective projection are extensively explored. They present a least-squares method to estimate the homology, initialized by four bi-tangent points, in which the homology is decomposed in terms of the central-axis projection and a vanishing point. Under-

standing the relationship between the central-axis projection, planar harmonic homology, and camera projection matrix, they provide a calibration procedure that operates over two SoR views (with visible bi-tangent points). Most importantly, Wong et al. (2002) presents a means to reconstruct the SoR shape from an apparent contour. This provides the second direction of the SoR shape-contour bijection.

Along similar lines, Colombo et al. (2002) provides a single-view SoR calibration algorithm that assumes a manually selected apparent contour and two ellipses. Another SoR reconstruction algorithm is built on top of this calibration algorithm in Colombo et al. (2004, 2005, 2006) that operates almost exclusively in projective space.

All previous algorithms require the camera calibration to compute the SoR reconstruction. Utcke and Zisserman (2003) sidesteps this requirement using two visible cross sections with the caveat that the resulting solution is up to a 2-DoF projective family, parameterized as a function of the camera calibration.

2.5. Summary

All the approaches presented in this chapter for the tasks of SoR pose and shape recovery make several limiting assumptions that are overcome by the methods presented in this thesis. The requirements of special points, cross sections, or pseudo-symmetry for 2D central-axis projection recovery are removed in Chapter 6 via the use of a two-point minimal problem for symmetric pose recovery. In Chapters 6 and 8, the requirements of visible cross sections and a known SoR shape model are removed for 3D central-axis recovery using two views for axis triangulation and structure from motion. Chapter 7 describes a method that recovers the pose of an SoR of a specified shape without visible cross sections. The various assumptions of prior work are summarized chronologically in Table 2.1, showing the progression of the techniques. All such techniques assume readily obtainable apparent contours, an assumption that is addressed in Chapters 9 and 10. These chapters frame SoR apparent contour selection and shape recovery as a perceptual grouping problem that is

	Ponce et al. (1989)	Dhome et al. (1990)	Glachet et al. (1991)	Glachet et al. (1992)	Forsyth et al. (1992)*	Sato and Binford (1993) *	Wong et al. (2002)*	Colombo et al. (2002)*	Utcke and Zisserman (2003)	Phillips et al. (2015)	Phillips et al. (2016)	Phillips and D. (2106) Abs	Phillips and D. (2106) SfM
Assumptions													
CAD Model		✓											
Generatrix				✓								✓	
Special points	✓	✓			✓		✓	✓					
1 cross section		✓		✓			✓						
2 cross sections						✓		✓	✓				
Near-orthography	✓		✓	✓		✓							
Unit aspect ratio							✓						
Intrinsics		✓		✓		✓				✓	✓	✓	✓
Stereo extrinsics										✓			
Single view	✓	✓	✓	✓	✓	✓						✓	
Stereo view										✓	✓		✓
Output													
2D-axis	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
3D-axis		✓		✓		✓	✓	✓		✓	✓	✓	✓
3D object center		✓		✓		✓	✓	✓		✓	✓	✓	✓
Intrinsics							✓	✓					
Projective shape					✓				✓				
Metric shape						✓	✓	✓		✓	✓		✓

*First paper of a highly related series

Forsyth et al. (1992); Zisserman et al. (1995)

Sato and Binford (1993); Gross and Boulton (1996)

Wong et al. (2002, 2003)

Colombo et al. (2002, 2004, 2005, 2006)

Table 2.1: Summary of approach assumptions and results, arranged chronologically. Prior to 2002, a near-orthographic camera was assumed to provide metric shape and pose recovery. This restriction was removed with the formulation of the 4-DoF central-axis homology under perspective projection and the addition of specialized SoR contour generator constraints. All prior work requires special points or cross sections to perform this task. The work presented in this thesis is not dependent on these features.

informed by the geometric constraints of SoRs, assuming known central-axes.

Part II

Preliminaries

3

Image formation and projection surfaces

In order for the light to shine so brightly, the darkness
must be present.

—Francis Bacon

Image formation begins in darkness, with a projection surface waiting to be struck with rays of light; rays that carry color and intensity information from an external scene through a small opening. This chapter describes the image formation process, introducing the prerequisite notation, models, and concepts for further exploration of SoR image profiles.

3.1. Notation and conventions

In this paper, points, vectors, normals, and lines are represented with bold lowercase letters, such as \mathbf{x} . They are expressed as column vectors, e.g., $\mathbf{x} = (x, y, z)^T$, with italics used for scalars. Matrices are represented with bold uppercase letters, such as \mathbf{M} . All symbols may have subscripts or superscripts to identify the entity, indicate the coordinate system, or denote the index. A left superscript always indicates the coordinate system in which an entity is expressed. Matrices that represent a mapping between coordinate systems have the source frame as a right subscript and the destination frame as a left superscript. Square brackets with a subscript may be used to indicate index. For example,

$${}^b\mathbf{x}_e = [{}^b\mathbf{R}_a]_i {}^a\mathbf{x}_e$$

shows that vector e expressed in coordinate system a is rotated to coordinate system b by the i^{th} rotation matrix from a to b .

A summary of the general conventions and symbol definitions is presented in Table 3.1.

Term type	Example	Note
Point / Vector	$\mathbf{x}_e, {}^a\mathbf{x}_e$	$(\{3, 4\} \times 1)$ Point e (in coordinate system a)
	$\mathbf{n}_e, {}^a\mathbf{n}_e$	(3×1) Normal e vector (in coordinate system a)
	$\mathbf{t}_b, {}^a\mathbf{t}_b$	(3×1) Origin (translation) of system a expressed in b
	\mathbf{t}_o	Camera center in object coordinate system
	\mathbf{t}_c	Object center in camera coordinate system
	$\hat{\mathbf{z}}$	The SoR central axis
Line	${}^a\mathbf{l}_e$	(3×1) Homogeneous line e in coordinate system a
Matrix	${}^b\mathbf{F}_a$	(4×4) Euclidean transform from system a to b
	${}^b\mathbf{R}_a$	(3×3) Euclidean rotation from system a to b
	$\mathbf{R}_z(\phi)$	(3×3) Euclidean rotation of angle ϕ about the z -axis
	${}^b\mathbf{H}_a$	(3×3) Homography from plane a to b
	\mathbf{K}	3×3 Projection from calibrated to image coordinates
	\mathbf{R}_ω	3×3 Infinitesimal rotation matrix
Plane	π_t	Tangent plane at an SoR surface point
	π_z	Plane through \mathbf{t}_o and $\hat{\mathbf{z}}$
	π_m	Plane through an SoR meridian
Matrix	\mathbf{S}_y	Reflection matrix through y -axis
Variables	\hat{d}, \hat{h}	Camera depth and height in canonical pose
	$\Delta\hat{d}, \Delta\hat{h}$	Camera pair baseline in term of depth and height
Function	$r(h)$	SoR generatrix radius (scaling) at height h
	$\rho(h)$	SoR generatrix first derivative at height h
Relation	$A \sim B$	Equivalence relation, A is equivalent to B by the relation
	$A \leftrightarrow B$	A corresponds with B
	$A \Leftrightarrow B$	Bi-conditional, A if and only B
	$\mathbf{x}_a \sim \mathbf{x}_b$	Projective equivalence relation, $\mathbf{x}_a = \lambda\mathbf{x}_b, \lambda \in \mathbb{R}_{\neq 0}$
Point sets	S	Object surface
	\mathcal{M}_θ	SoR Meridian at azimuth θ in object cylindrical coordinates
	\mathcal{P}_h	SoR Parallel at height h in object cylindrical coordinates
Scalars	ϕ	Pitch attitude
	θ	Azimuth angle

Table 3.1: Summary of mathematical terms and conventions. This table summarizes specific mathematical terms and conventions that are used throughout this thesis.

3.2. Image formation

Image formation as a physical process is a complex interaction of geometry and optics. Indeed, fully and accurately encoding these interactions towards the goal of photo-realistic rendering methods and techniques is an active and thriving area of research. This section describes the aspects of the image formation process that are modeled and utilized throughout this thesis.

3.2.1. Perspective projection and the pinhole camera model

The term “camera” in English comes from “camera obscura”, Latin for “dark room”. A camera obscura is a dark room or box with a small hole on one side. Light passes through the hole to produce an image of the outside scene on the opposing wall. This concept was adapted in the form of a “pinhole” camera that was prevalent in early photography. The pinhole opening in the camera acts as an aperture, a center through which light from the scene geometry passes. These rays of light then strike a film plate that captures the image.

The pinhole camera model is used as the mathematical basis describing image formation under *perspective projection*. As a mathematical model, the *center of projection* and 2D *image plane* are conceptual analogues to the pinhole and the film plate, respectively. Light rays from scene objects are modeled as 3D points that undergo a central projection to the image plane.

3.2.2. Calibrated image plane

The camera coordinate system uses the center of projection as the origin with the positive z -axis representing the normal of the image plane. Mathematically, all 3D points are mapped to the image plane through the central projection. In practice, the only points that are considered are those in front of the camera with a positive z -coordinate. The image plane is taken as the xy plane at $z = 1$, with the top of the image in the negative y direction (see Figure 3.1).

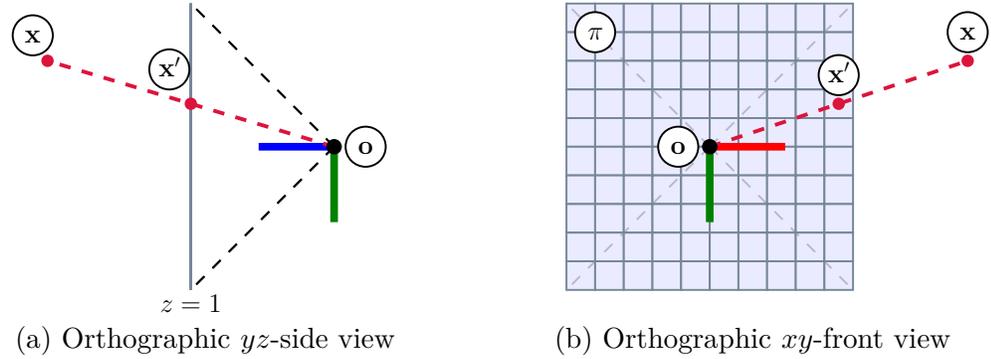


Figure 3.1: Calibrated camera coordinate system. From the orthographic side view (a), the 3D point \mathbf{x} is projected towards the projection center \mathbf{o} to create 2D image point \mathbf{x}' on the image plane at $z = 1$. The dashed black lines represent the camera's view frustum. In the orthographic front view (b), the 2D location of \mathbf{x}' on the image plane π is seen. The green y -axis points down, the red x -axis points to the right, and the blue z -axis points forward from the projection center towards the image plane.

This representation is considered the *standard* or *natural camera coordinate system*. It is also called the *calibrated camera coordinate system*, as it abstracts away intrinsic properties of the physical imaging device, which must be accounted for through system calibration. In an uncalibrated system the image plane is represented in *image coordinates*, which is typically expressed in terms of *pixels*, whereas the calibrated image plane is typically expressed in a real-world metric unit, such as meters.

3.2.3. Coordinate transform

It is useful to define (at least) two coordinate systems, one for describing positions relative to the scene geometry, and the other relative to the camera.

A 3D point ${}^o\mathbf{x}$ in object coordinates is transformed to camera point ${}^c\mathbf{x}$ by the 3×3 rotation matrix ${}^c\mathbf{R}_o$ and object center ${}^c\mathbf{t}_o$, expressed in the camera coordinate system, by

$${}^c\mathbf{x} = {}^c\mathbf{R}_o {}^o\mathbf{x} + {}^c\mathbf{t}_o \quad (3.1)$$

or homogeneously by

$$\begin{pmatrix} {}^c\mathbf{x} \\ 1 \end{pmatrix} = \underbrace{\begin{pmatrix} {}^c\mathbf{R}_o & {}^c\mathbf{t}_o \\ \mathbf{0} & 1 \end{pmatrix}}_{{}^c\mathbf{F}_o} \begin{pmatrix} {}^o\mathbf{x} \\ 1 \end{pmatrix}, \quad (3.2)$$

where ${}^c\mathbf{F}_o$ is the homogeneous transform from the object to the camera system.

3.2.4. Perspective projection

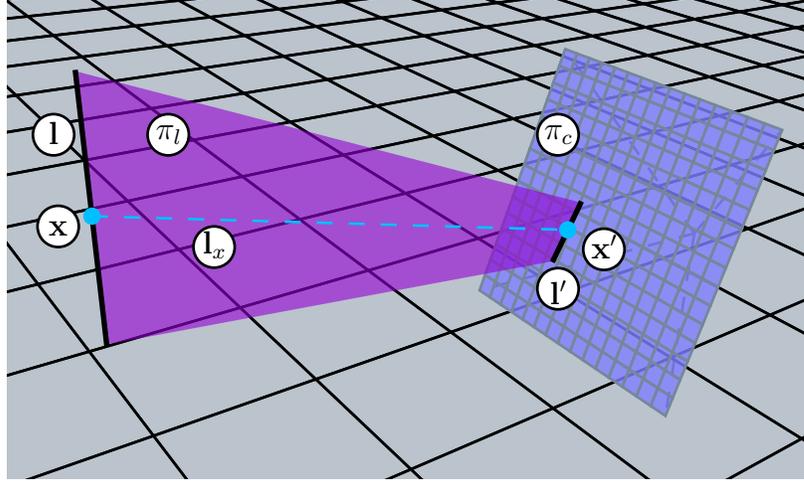


Figure 3.2: The back-projection of points and lines. The 3D line l and point x project to image plane π_c as 2D line l' and x' . The back-projection of l' forms the plane π_l , while the 2D point x' back-projects to 3D line l_x .

Central projection of a camera point ${}^c\mathbf{x} = (x, y, z)^\top$ to the x - y image plane at $z = 1$ is accomplished by *projective division* as

$${}^c\mathbf{x}' = (x/z, y/z, z/z)^\top = (x/z, y/z, 1)^\top, \quad (3.3)$$

where ${}^c\mathbf{x}'$ is said to be the homogeneous representation of the non-homogeneous 2D image point $(x/z, y/z)^\top$.

Observe that any point of the form $(\lambda u, \lambda v, \lambda)^\top$ maps to the same point $(u, v, 1)^\top$ under

projective division, yielding the projective equivalence relation

$$(x_a, y_a, z_a)^\top \sim (x_b, y_b, z_b)^\top \leftrightarrow (\lambda(x_a, y_a, z_a)^\top = (x_b, y_b, z_b)^\top, \lambda \in \mathbb{R}_{\neq 0}). \quad (3.4)$$

Homogeneous coordinates are a convenient means of representing an equivalence class of geometric entities under projective division. The equivalent class elements of a homogeneous point span a 3D line in space, while those of a homogeneous line span a 3D plane (see Figure 3.2). The pre-image of a homogeneous entity under projective division is called its back-projection. These concepts are vital to understanding the interaction between 3D surfaces and their images.

3.2.5. Homogeneous image lines and 3D planes

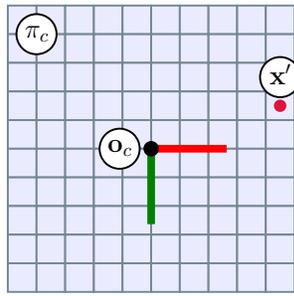
Image lines can also be represented homogeneously by a 3D vector in a manner identical to image points, however its geometric interpretation might not be as readily apparent. An image line can be viewed as a collection of image points, the back-projection of which is a collection of coplanar rays. When expressed in calibrated coordinates, these coplanar rays sweep out a plane (Figure 3.2) with normal \mathbf{n} that passes through the camera center. The homogeneous form of an image line is precisely \mathbf{n} , the normal of the 3D plane it back-projects to. A homogeneous image point \mathbf{x} is on the image line \mathbf{n} if the 3D ray $\lambda\mathbf{x}$ is contained within the plane. To express this mathematically, it is enough to verify that the 3D ray is perpendicular to the plane normal, satisfying

$$\mathbf{n}^\top \mathbf{x} = 0. \quad (3.5)$$

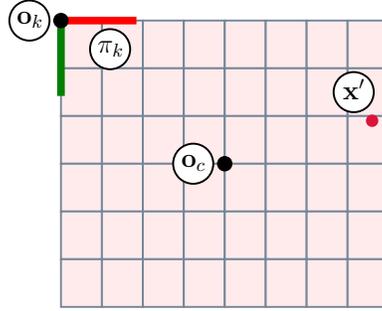
Note that both \mathbf{n} and \mathbf{x} can be arbitrarily scaled by λ without violating this equation.

3.2.6. Uncalibrated image plane

The calibrated image plane represents the 2D projection of the 3D world points in a metric unit such as meters, with the center of projection at the origin. These properties allow it to



(a) Calibrated image plane



(b) Uncalibrated image plane

Figure 3.3: Uncalibrated camera coordinate system. The calibrated image plane π_c (a) is typically represented with metric real-world units such as meters, with the projection center at the origin \mathbf{o}_c . The uncalibrated image plane π_k (b) is typically expressed in pixels, which may not be square. As such the image plane is stretched unequally in the x and y directions from the calibrated image plane. The projection center \mathbf{o}_c , also called principal point, usually does not coincide with the image plane origin \mathbf{o}_k . As a result of these differences, the same image point \mathbf{x}' may be represented on both image planes by drastically different coordinates.

be used directly when reasoning in 3D Euclidean space. In practice, captured image points are typically represented in a different coordinate system that mirrors the physical geometry of the imaging device. This system is typically called *image coordinates* and is represented in terms of pixels. The uncalibrated image plane is commonly related to the calibrated image plane by an invertible 3×3 *intrinsic matrix* \mathbf{K} with four degrees of freedom of the form

$$\mathbf{K} = \begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix}.$$

The 2D point $(c_x, c_y)^\top$ represents the image location of the projection center, also called the principle point, in the image plane. When this point is non-zero, it implies that the uncalibrated image plane origin is shifted from that of the calibrated image plane (see Figure 3.3). The parameters f_x and f_y encode the focal length and metric pixel dimensions. If f_x and f_y are equal, then the pixels are square and the camera is said to have *unit aspect ratio*. Otherwise, the pixels are non-square and the image plane is stretched in the x and y directions by different amounts.

\mathbf{K} and its inverse \mathbf{K}^{-1} allow points and geometric entities to be transformed between the calibrated and uncalibrated coordinate systems. A non-homogeneous calibrated point is transformed by

$${}^k\mathbf{x} \sim \mathbf{K} [{}^c\mathbf{R}_o {}^o\mathbf{x} + {}^c\mathbf{t}_o]. \quad (3.6)$$

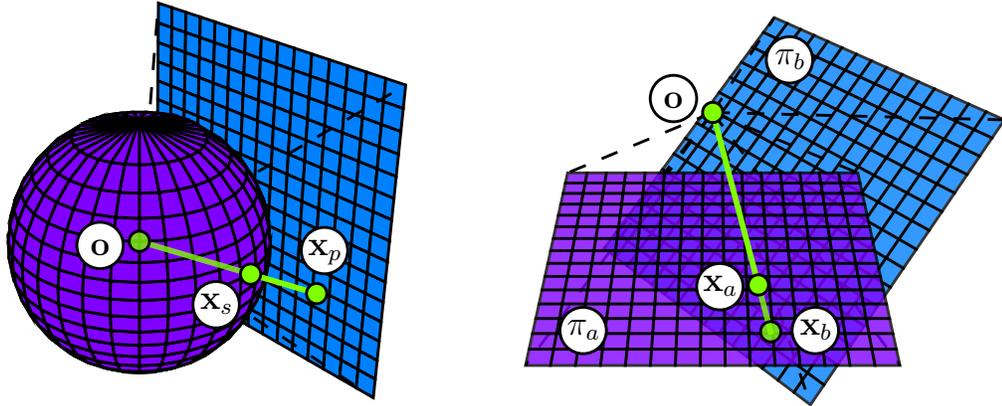
Using a homogeneous representation, the 3×4 form of \mathbf{K} simultaneously transforms to image coordinates and produces a drop in dimensionality with

$${}^k\mathbf{x} \sim \begin{pmatrix} \mathbf{K} & \mathbf{0} \end{pmatrix} \begin{pmatrix} {}^c\mathbf{R}_o & {}^c\mathbf{t}_o \\ \mathbf{0} & 1 \end{pmatrix} \begin{pmatrix} {}^o\mathbf{x} \\ 1 \end{pmatrix}. \quad (3.7)$$

3.3. Projection surface bijections

In object coordinates, the location of the projection center is represented as a 3×1 translation column vector ${}^o\mathbf{t}_c$. The orientation of the *image plane* of the calibrated camera about ${}^o\mathbf{t}_c$ is specified by the rotation ${}^c\mathbf{R}_o$. All rays through this center of projection also pass through a unit sphere centered at ${}^o\mathbf{t}_c$. This unit sphere is called the *image sphere*, as the projection of the 3D scene points onto this sphere encodes all the information necessary for image formation. Modulo field of view, this representation contains no more or less information than projection onto an image plane.

Representing an image using different surfaces of projection may be desirable as each projection surface exhibits its own set of geometric properties. For example, the image sphere is useful for the treatment of any projective geometry that is invariant to the specific choice of image plane, i.e., invariant to the camera orientation. The tangency between camera rays and scene surfaces is such an invariant property, as introduced in Section 4.1. It is this property that allows for seamless panoramic stitching of planar images that are taken with the same (approximate) projection center, as well as the synthesis of arbitrary image plane views from such a panorama.



(a) Plane-sphere bijection

(b) Plane-plane bijection

Figure 3.4: Projective bijections through shared projection center \mathbf{o} . (a) shows the bijection between the point \mathbf{x}_s on the image sphere and \mathbf{x}_p on the image plane. (b) shows the bijection between the point \mathbf{x}_a on plane π_a and \mathbf{x}_b on plane π_b .

Panoramic images are an ideal example for discussing three important projection surface bijections that are utilized in this thesis. The plane-sphere bijection maps images to the image sphere, where the image rays from overlapping view frustums blend seamlessly to create an omni-directional image on the sphere that may be back-projected onto any image plane of the same projection center. The plane-plane bijection maps between image planes directly, skipping the intermediate view sphere projections for the synthesis of an arbitrary image plane view. These two bijections are shown in Figure 3.4. Finally, the sphere-cylinder bijection is commonly used to store panoramic images as one continuous “flat” image.

3.3.1. Plane-sphere bijection

A point ${}^o\mathbf{x}$ on the image sphere in the object coordinate system is expressed in the camera coordinate system as ${}^c\mathbf{x} = {}^c\mathbf{R}_o {}^o\mathbf{x} + {}^c\mathbf{t}_o = (x, y, z)^\top$, where ${}^c\mathbf{t}_o$ is the object origin expressed in camera coordinates.

There is a bijective function $f : \mathcal{S} \rightarrow \mathcal{P}$ between image sphere and image plane points,

where

$$\mathcal{S} = \{\mathbf{x} \in \mathbb{R}^3 | \mathbf{x} = (x, y, z)^\top, \|\mathbf{x}\| = 1\}$$

is the set of unit sphere points and

$$\mathcal{P} = \{\mathbf{x} \in \mathbb{R}^3 | \mathbf{x} = (x, y, z)^\top, z = 1\}$$

is the set of image plane points.

The function

$$f(\mathbf{x}) = (x, y, z)^\top / z$$

projects the unit image sphere point to the image plane in the standard projective manner.

The inverse function

$$f^{-1}(\mathbf{x}) = \frac{\mathbf{x}}{\|\mathbf{x}\|}$$

projects the image plane point to the unit sphere by dividing by its magnitude.

By verifying $f(f^{-1}(\mathbf{x})) = \mathbf{x}, \forall \mathbf{x} \in \mathcal{P}$:

$$\begin{aligned} f(f^{-1}(\mathbf{x})) &= \mathbf{x} \\ f(f^{-1}((x, y, 1)^\top)) &= (x, y, 1)^\top \\ f\left(\frac{(x, y, 1)^\top}{\|(x, y, 1)^\top\|}\right) &= (x, y, 1)^\top \\ \frac{(x, y, 1)^\top \|(x, y, 1)^\top\|}{\|(x, y, 1)^\top\|} &= (x, y, 1)^\top \\ (x, y, 1)^\top &= (x, y, 1)^\top \end{aligned}$$

and $f^{-1}(f(\mathbf{x})) = \mathbf{x}$, $\forall \mathbf{x} \in \mathcal{S}$:

$$\begin{aligned}
f^{-1}(f(\mathbf{x})) &= \mathbf{x} \\
f^{-1}\left(f\left((x, y, z)^\top\right)\right) &= (x, y, z)^\top \\
f^{-1}\left(\left(x/z, y/z, 1\right)^\top\right) &= (x, y, z)^\top \\
\frac{(x/z, y/z, 1)^\top}{\|(x/z, y/z, 1)^\top\|} &= (x, y, z)^\top \\
\frac{z(x/z, y/z, 1)^\top}{\|(x, y, z)^\top\|} &= (x, y, z)^\top \\
\frac{z(x/z, y/z, 1)^\top}{1} &= (x, y, z)^\top \\
(x, y, z)^\top &= (x, y, z)^\top
\end{aligned}$$

it is shown that f and f^{-1} are indeed each other's inverse, implying f is invertible and consequently a bijection. This bijection is illustrated in Figure 3.4a.

3.3.2. Plane-plane bijection (two-plane homography)

If a 3D point ${}^a\mathbf{x}$ in camera a is related to a 3D point ${}^b\mathbf{x}$ in camera b by

$${}^b\mathbf{x} = {}^b\mathbf{R}_a {}^a\mathbf{x} + {}^b\mathbf{t}_a,$$

then the homology from the image plane of camera a to the image plane of b is

$${}^b\mathbf{H}_a = {}^b\mathbf{R}_a + \begin{pmatrix} \mathbf{0} & \mathbf{0} & {}^b\mathbf{t}_a \end{pmatrix}. \quad (3.8)$$

If the two cameras share the same center of projection, ${}^b\mathbf{t}_a = \mathbf{0}$ then ${}^b\mathbf{H}_a$ is simply the rotation matrix ${}^b\mathbf{R}_a$ relating the two cameras.

This result is seen by first observing that the image of a point ${}^a\mathbf{x}$ in camera a is transferred to ${}^b\mathbf{x}'$ on camera b 's image plane by the following steps:

1. Project ${}^a\mathbf{x}$ to image plane π_a by projective division as ${}^a\mathbf{x}' = {}^a\mathbf{x}/z_a$.

2. Represent ${}^a\mathbf{x}'$ in camera b coordinates by ${}^b\mathbf{x} = {}^b\mathbf{R}_a {}^a\mathbf{x}' + {}^b\mathbf{t}_a$.
3. Project ${}^b\mathbf{x}$ to ${}^b\mathbf{x}'$ by projective division.

In homogeneous coordinates, the relationship between ${}^a\mathbf{x}$ and ${}^b\mathbf{x}'$ becomes

$${}^b\mathbf{x}' \sim {}^b\mathbf{R}_a ({}^a\mathbf{x}/z_a) + {}^b\mathbf{t}_a. \quad (3.9)$$

This is massaged into the form of a homography as follows:

$$\begin{aligned} {}^b\mathbf{x}' &\sim {}^b\mathbf{R}_a ({}^a\mathbf{x}/z_a) + {}^b\mathbf{t}_a && \text{(Eq. 3.9)} \\ {}^b\mathbf{x}' &\sim {}^b\mathbf{R}_a {}^a\mathbf{x} + z_a {}^b\mathbf{t}_a && \text{(By projective equivalence)} \\ {}^b\mathbf{x}' &\sim x_a \mathbf{r}_x + y_a \mathbf{r}_y + z_a \mathbf{r}_z + z_a {}^b\mathbf{t}_a && \text{(By matrix expansion)} \\ {}^b\mathbf{x}' &\sim x_a \mathbf{r}_x + y_a \mathbf{r}_y + z_a (\mathbf{r}_z + {}^b\mathbf{t}_a) && \text{(By additive distribution)} \\ {}^b\mathbf{x}' &\sim \begin{pmatrix} \mathbf{r}_x & \mathbf{r}_y & (\mathbf{r}_z + {}^b\mathbf{t}_a) \end{pmatrix} {}^a\mathbf{x} && \text{(By matrix construction)} \\ {}^b\mathbf{x}' &\sim \left({}^b\mathbf{R}_a + \begin{pmatrix} \mathbf{0} & \mathbf{0} & {}^b\mathbf{t}_a \end{pmatrix} \right) {}^a\mathbf{x} && \text{(By matrix additivity)} \\ {}^b\mathbf{x} &\sim {}^b\mathbf{H}_a {}^a\mathbf{x}. && \text{(Desired result, eq. 3.8)} \end{aligned}$$

If the two cameras share a center of projection, then ${}^b\mathbf{t}_a = \mathbf{0}$, and the homography reduces to ${}^b\mathbf{R}_a$. This homography is illustrated in Figure 3.4b.

4

Contour generators and surfaces of revolution

We are to admit no more causes of natural things than such as are both true and sufficient to explain their appearances.

—Isaac Newton

When an image is formed of a solid 3D object, a dimension reduction occurs as the object's 3D surface is represented as a 2D surface in the image. Accordingly, the 3D surface boundary generally projects to some 2D image boundary, outlining the surface profile or silhouette. This chapter begins by summarizing the work of Giblin and Weiss (1987), who first explored occluding contours and the wealth of surface information they contain in their paper entitled *Reconstruction of Surfaces from Profiles*.

Specializing these results for surfaces of revolution, Glachet et al. (1992) provide an SoR occluding contour projective rendering equation given the generatrix and pose. Wong et al. (2002) then provide the projective constraints for metric generatrix reconstruction given the occluding contour and pose. The second half of this chapter presents this bijection in a unified framework, providing a Euclidean decomposition of the SoR occluding contour forward-projection equation, parameterized by pose. This framework enables the formulation of the minimal problems for pose and generatrix recovery in subsequent chapters. It also provides a means to aggregate evidence from multiple, potentially noisy, images for generatrix reconstruction and apparent contour perceptual grouping.

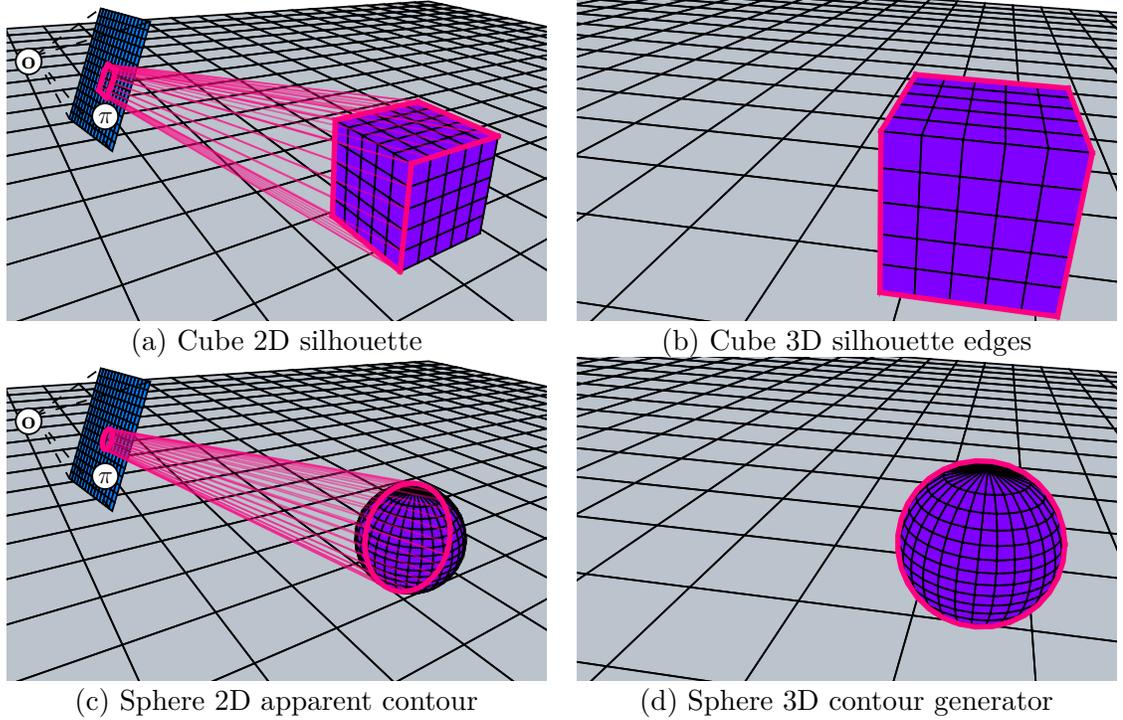


Figure 4.1: 2D and 3D boundaries of smooth and discontinuous surfaces. The 2D and 3D boundaries are outlined in thick red lines. In (a) and (c), the 3D boundary boundaries are shown for a cube and sphere, respectively. The fine red lines represent the 3D boundary projection onto image plane π through projection center \mathbf{o} . The resulting profiles are shown in (b) and (d), respectively. Note that the sphere has a smooth 3D boundary, while the cube’s 3D boundary is composed entirely of surface discontinuities.

4.1. General contour generators

The 2D boundary of an imaged surface, called the *apparent contour*, represents the projection of some 3D surface boundary, called the *extremal boundary* or *contour generator*, that divides the object into visible and occluded parts. A subtle yet important distinction is made between boundary points that correspond to smooth surface geometry versus discontinuous regions such as edges and corners. Discontinuous contour generator points are appropriately called surface *edges*, while smooth regions are called surface *limbs*. *Occluding contours* are the smooth apparent contours resulting from the projection of surface limbs. Figure 4.1 shows the apparent contour of two surfaces, a cube and a sphere.

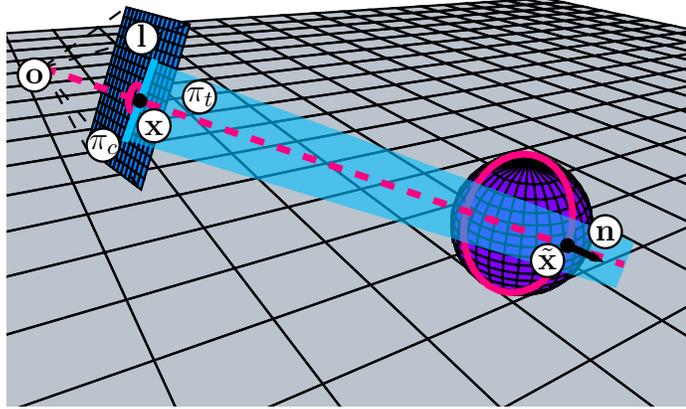


Figure 4.2: General contour generator relationships. The 3D point $\tilde{\mathbf{x}}$ with normal \mathbf{n} is part of the contour generator (solid red circle), and therefore its tangent plane π_t includes the image ray through point \mathbf{x} on image plane π_c with projection center \mathbf{o} . The tangent plane projects to homogeneous line \mathbf{l} , with a representation equal to \mathbf{n} when expressed in calibrated coordinates.

For surfaces such as the cube with no limbs, the apparent contour is composed entirely of projected edges. For smooth surfaces however, the relationship between the 2D and 3D surface boundaries is a complex function of the projection center and surface shape.

For a smooth surface \mathcal{S} with well-defined surface normals at each point, it is possible to define the contour generator that produces the occluding contour for a specific camera center ${}^o\mathbf{t}_c$. It is important to note that the choice of ${}^o\mathbf{t}_c$ fully defines the superset of image rays, a subset of which intersects the contour generator. The contour generator is composed of those image rays that intersect the surface at a single point ${}^o\mathbf{x}$ and lay within the tangent plane defined at ${}^o\mathbf{x}$. Equivalently, any point ${}^o\mathbf{x}$ whose tangent plane contains the camera center ${}^o\mathbf{t}_c$ belongs to the contour generator. These relationships are demonstrated in Figure 4.2.

Explicitly, a surface point ${}^o\mathbf{x}$ with normal ${}^o\mathbf{n}$ is part of a contour generator for camera center ${}^o\mathbf{t}_c$ if for some image ray ${}^o\mathbf{v}$ the following equations hold:

$$\lambda {}^o\mathbf{v} + {}^o\mathbf{t}_c = {}^o\mathbf{x} \quad (\text{ray-surface intersection constraint}) \quad (4.1)$$

$${}^o\mathbf{n}^\top({}^o\mathbf{x} - {}^o\mathbf{t}_c) = 0 \text{ (ray-surface tangency constraint)}. \quad (4.2)$$

These are the central constraints of Giblin and Weiss (1987) that enable reasoning about the profiles of arbitrary smooth geometry surfaces.

4.1.1. Apparent contours and projection surfaces

It is important to reiterate that the contour generators are solely dependent on the choice of surface S and projection center ${}^o\mathbf{t}_c$. The effect of this is that their corresponding apparent contours are a sole function of the specific choice of the projection surface. That is to say, the different apparent contours formed by a central projection onto the image sphere, standard camera image plane, or uncalibrated image plane, all originate from identical contour generators. Additionally, as discussed in Section 3.3, there are projective bijections between images planes and image spheres with the same center of projection that allow geometric reasoning to be performed on whichever surface is most mathematically convenient.

4.2. Surface of revolution representation

Viewed as a specialization of SHGCs, surfaces of revolutions are formed by sweeping a circular cross section along a central axis, $\hat{\mathbf{z}}$, while the circle is scaled by scaling function $r(z)$. Alternatively, the function $r(z)$ can be treated as a planar curve, called a *generatrix*, that is revolved around $\hat{\mathbf{z}}$, sweeping out the surface of the SoR. Figure 4.3 presents this sweeping process under both interpretations.

Cylindrical coordinates (r, θ, h) are a natural way of expressing SoR points and normals, with the cylindrical height extending along an upward pointing central axis of revolution, $\hat{\mathbf{z}}$. For the purpose of relating the SoR pose to the camera pose, the positive x -axis is treated as the “front” of the SoR. The standard camera coordinate system is used, with the positive z -axis aligned with the optical axis, and the x -axis pointing to the camera’s right. The relationship between these two coordinate systems are illustrated in Figure 4.4.

The generatrix function $r(h)$ determines the radius for a given height and has a correspond-



(a) Planar generatrix curve (b) Scaled cross sections (c) Generatrix rotations

Figure 4.3: Surface of revolution formation. The planar generatrix curve in (a) is used to produce the SoR by sweeping through the space in two different ways. In (b) it is used to scale the circular cross section that is swept along $\hat{\mathbf{z}}$, while in (c) it is swept in revolution about $\hat{\mathbf{z}}$.

ing derivative function $\rho(h)$ that yields the derivative $\frac{dr(h)}{dh}$. Using these two functions, a SoR surface point $\mathbf{x}_o \stackrel{\text{def}}{=} (x_o, y_o, h)^\top$ and corresponding normal $\mathbf{n}_o \stackrel{\text{def}}{=} (a_o, b_o, c_o)$ are parameterized by height h and azimuth angle θ in the SoR coordinate system as

$$\mathbf{x}_o(\theta, h) = (r(h) \cos \theta, r(h) \sin \theta, h)^\top, \quad (4.3)$$

and

$$\mathbf{n}_o(\theta, h) \propto (\cos \theta, \sin \theta, -\rho(h))^\top \propto (x_o, y_o, -r(h)\rho)^\top \quad (\text{see Section 4.2.2}), \quad (4.4)$$

respectively. Notice that the x and y components of the surface normal are proportional to those of the surface point. This property proves useful for constraining pose and shape recovery from the occluding contour.

4.2.1. Surface point representation

The SoR surface point set \mathcal{S} is generated by rotating the generatrix $r(h)$ about the z -axis, with

$$\mathcal{S} = \{\mathbf{x}_o(\theta, h) \mid 0 \leq \theta \leq 2\pi, h \in \text{dom}(r)\}. \quad (4.5)$$

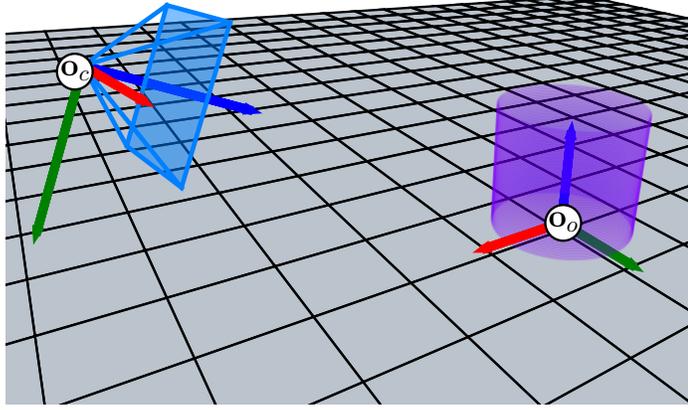


Figure 4.4: Object and camera coordinate systems. The origins of the object and camera coordinate systems marked with \mathbf{o}_o and \mathbf{o}_c , respectively. The x , y , and z -axes are represented by the red, green, and blue arrows, respectively. The z -axis is treated as the up or height axis for the object coordinate system. The camera points in the z -axis direction, with the positive y -axis pointing towards the bottom of the image and the positive x -axis pointing to the right.

A direct consequence of this is that \mathcal{S} maps onto itself under any rotation $\mathbf{R}_z(\phi)$ of angle ϕ about the z -axis. This can be seen by applying an arbitrary rotation about the z -axis to the definition of the SoR point set:

$$\mathbf{R}_z(\phi)\mathcal{S} = \mathbf{R}_z(\phi)\{\mathbf{x}_o(\theta, h) \mid 0 \leq \theta \leq 2\pi, h \in \text{dom}(\mathbf{r})\}$$

$$\mathbf{R}_z(\phi)\mathcal{S} = \{\mathbf{R}_z(\phi)\mathbf{x}_o(\theta, h) \mid 0 \leq \theta \leq 2\pi, h \in \text{dom}(\mathbf{r})\}$$

$$\mathbf{R}_z(\phi)\mathcal{S} = \{\mathbf{x}_o(\theta + \phi, h) \mid 0 \leq \theta \leq 2\pi, h \in \text{dom}(\mathbf{r})\}$$

$$\mathbf{R}_z(\phi)\mathcal{S} = \{\mathbf{x}_o(\theta, h) \mid \phi \leq \theta \leq 2\pi + \phi, h \in \text{dom}(\mathbf{r})\}$$

$$\mathbf{R}_z(\phi)\mathcal{S} = \{\mathbf{x}_o(\theta, h) \mid 0 \leq \theta \leq 2\pi, h \in \text{dom}(\mathbf{r})\}$$

$$\mathbf{R}_z(\phi)\mathcal{S} = \mathcal{S}.$$

This relationship is commonly used to reduce the number of parameters that are needed to express the SoR's translation from three to two (Glachet et al., 1992; Wong et al., 2002).

4.2.2. Surface normal representation

The surface of an SoR contains points $\mathbf{x}(\theta, h)$ parameterized by azimuth angle θ and height h . The surface normal is orthogonal to both partial vector derivatives of $\mathbf{x}(\theta, h)$ with respect to θ and h . Therefore, the surface normal is computed using the cross product as follows:

$$\mathbf{n}_o(\theta, h) = \frac{\partial \mathbf{x}_o(\theta, h)}{\partial \theta} \times \frac{\partial \mathbf{x}_o(\theta, h)}{\partial h} \quad (\text{Step 1})$$

$$\mathbf{n}_o(\theta, h) = \frac{\partial}{\partial \theta} (r(h) \cos \theta, r(h) \sin \theta, h)^\top \times \frac{\partial}{\partial h} (r(h) \cos \theta, r(h) \sin \theta, h)^\top \quad (\text{Step 2})$$

$$\mathbf{n}_o(\theta, h) = (-r(h) \sin \theta, r(h) \cos \theta, 0)^\top \times (\rho \cos \theta, \rho \sin \theta, 1)^\top \quad (\text{Step 3})$$

$$\mathbf{n}_o(\theta, h) = \begin{pmatrix} r(h) \cos \theta (1) - (0) \rho \sin \theta \\ (0) \rho \cos \theta - -r(h) \sin \theta (1) \\ -r(h) \sin \theta \rho \sin \theta - r(h) \cos \theta \rho \cos \theta \end{pmatrix} \quad (\text{Step 4})$$

$$\mathbf{n}_o(\theta, h) = \begin{pmatrix} r(h) \cos \theta \\ r(h) \sin \theta \\ -r(h) \rho (\cos^2 \theta + \sin^2 \theta) \end{pmatrix} \quad (\text{Step 5})$$

$$\mathbf{n}_o(\theta, h) = r(h) \begin{pmatrix} \cos \theta \\ \sin \theta \\ -\rho \end{pmatrix} \quad (\text{Step 6})$$

$$\mathbf{n}_o(\theta, h) \sim (\cos \theta, \sin \theta, -\rho)^\top. \quad (\text{Step 7})$$

Step 6 shows the normal for a point $\mathbf{x}_o = (x, y, h)$ on parallel \mathcal{P}_h as

$$\mathbf{n}_o(\theta, h) = r(h) \begin{pmatrix} \cos \theta \\ \sin \theta \\ -\rho \end{pmatrix} = \begin{pmatrix} r(h) \cos \theta \\ r(h) \sin \theta \\ -r(h) \rho \end{pmatrix} = \begin{pmatrix} x \\ y \\ -r(h) \rho \end{pmatrix}. \quad (4.6)$$

This form is first seen in Glachet et al. (1992), and is used to relate the contour generator of an SoR to its generatrix.

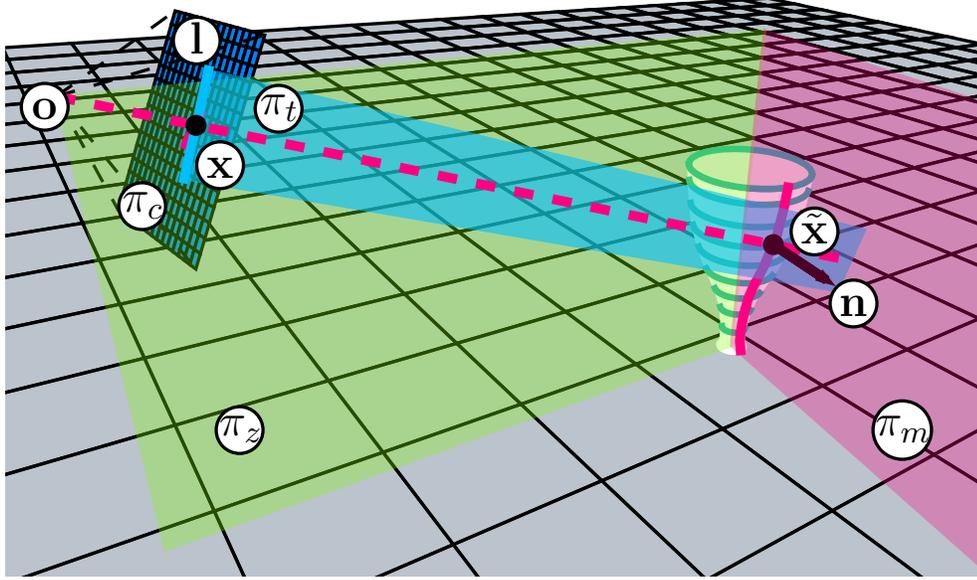


Figure 4.5: SoR contour generator constraints. As is the case with general contour generators, the tangent plane π_t with normal \mathbf{n} must contain the contour generator point $\tilde{\mathbf{x}}$, its projection \mathbf{x} on π_c , as well as the camera center \mathbf{o} . Additionally, $\tilde{\mathbf{x}}$ must be contained within the meridian plane π_m that contains \mathbf{n} .

4.3. Meridian and parallel constraints

Parameterized by azimuth θ and height h , an SoR can be viewed as the infinite composition of *parallels* of constant h , or alternatively as *meridians* of constant θ (see Figure 4.3). A parallel is the circle of surface points generated at a constant height, named for the fact that all such circles are parallel to each other, while perpendicular to the central axis $\hat{\mathbf{z}}$. Meridians are points of the same azimuth that are contained in the half plane π_m with normal \mathbf{m} that extends from $\hat{\mathbf{z}}$ outwards. The parallel \mathcal{P}_h is an infinite point set defined as

$$\mathcal{P}_h = \{\mathbf{x}_o | \mathbf{x}_o = (x, y, h) \wedge x^2 + y^2 = r(h)^2\}, \quad (4.7)$$

while the meridian \mathcal{M}_θ is defined as

$$\mathcal{M}_\theta = \{\mathbf{x}_o(\theta, h) | h \in \text{dom}(r)\}. \quad (4.8)$$

By the surface definition, an SoR surface point \mathbf{x}_o is constrained to belong to some parallel and to some meridian. This yields the surface-parallel intersection constraint

$$x_o^2 + y_o^2 = r(h)^2, \quad (4.9)$$

and the surface-meridian intersection constraint

$$\mathbf{x}_o^T \mathbf{m} = 0. \quad (4.10)$$

Additionally, the normal \mathbf{n}_o of any SoR surface point \mathbf{x}_o must also be contained in the same meridian, yielding the normal-meridian coplanarity constraint

$$(\hat{\mathbf{z}}_o \times \mathbf{n}_o)^T \mathbf{x}_o = 0. \quad (4.11)$$

All geometric constraints for SoR contour generators are shown in Figure 4.5. These constraints are encoded in the methods of Wong et al. (2002) and Colombo et al. (2005) to achieve matrix generatrix reconstruction.

4.4. Occluding contour and generatrix bijection

Recall that the contour generator of a surface is a function of the projection center and surface shape, and that the occluding contour of an SoR is the central projection of the 3D contour generator onto some surface of projection. In this section, the bijection between the occluding contour and generatrix is shown with a calibrated camera in a canonical pose. Combined with the projective bijections of Section 3.3 and the projection center equivalence relation of Section 4.4.2, this is sufficient to extend the bijection to a camera of arbitrary pose and 4-DoF projection matrix.

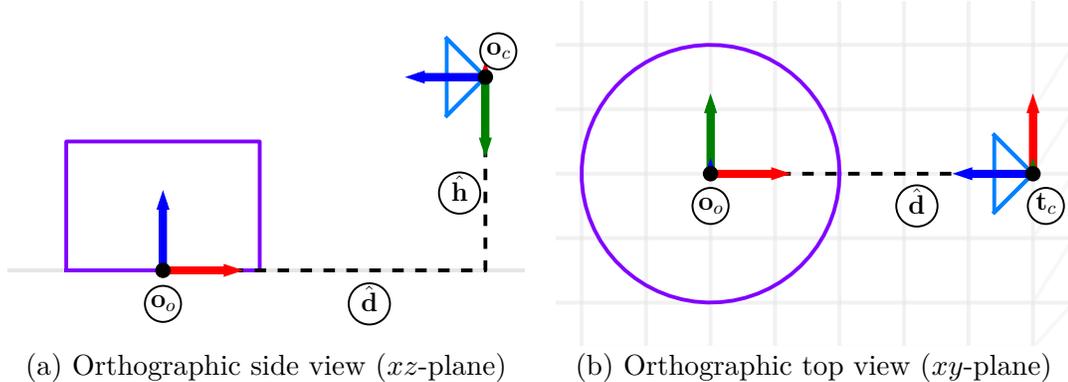


Figure 4.6: Canonical camera pose representation. As shown by the two orthographic views, the canonical camera pose places the camera center \mathbf{o}_c at height \hat{h} above the xy -plane and at a distance of \hat{d} along the xy -plane away from the object origin at \mathbf{o}_o . The normal of the image plane intersects the central axis perpendicularly. The result of this configuration is that any SoR appears symmetric in the canonical image.

4.4.1. Canonical camera

The canonical camera is defined as an upright calibrated camera in the object coordinate system, with its z -axis perpendicular to and intersecting the central axis $\hat{\mathbf{z}}$ of the SoR. It has a projection center of the form ${}^c\mathbf{t}_o = (\hat{d}, 0, \hat{h})^\top$ in the object coordinate system, where \hat{d} is the camera's distance from the SoR origin along the xy -plane, and \hat{h} is the camera height. The pose of the canonical camera with respect to the object frame is represented succinctly by the homogeneous transform

$${}^o\mathbf{F}_g = \begin{bmatrix} 0 & 0 & -1 & \hat{d} \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & \hat{h} \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (4.12)$$

which is depicted in Figure 4.6. Most reconstruction methods define such a canonical view (Wong et al., 2002; Utcke and Zisserman, 2003; Colombo et al., 2005) into which the input image is transformed as a penultimate step in generatrix reconstruction.

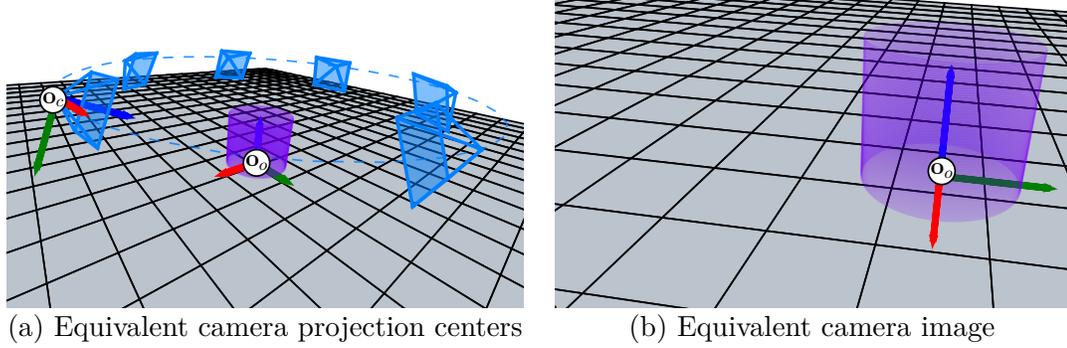


Figure 4.7: Projection center equivalence. (a) illustrates the set of camera poses that are equivalent due to the object’s symmetry about its z -axis. Rotating the camera around the z -axis of the static object is the same relative motion as rotating the rotationally symmetric object about its axis with the camera static. (b) shows the image seen from \mathbf{o}_t that has invariant apparent contours for all equivalent poses.

4.4.2. Center of projection equivalence relation

Surfaces of revolution have the convenient property of circular symmetry; the surface maps onto itself under rotation about its central axis $\hat{\mathbf{z}}$ by any arbitrary angle. This can be seen by examining the surface point set, as shown in Section 4.2.1. Due to this property, the same shape is seen by a static camera as an SoR rotates about $\hat{\mathbf{z}}$. Expressing this relative motion from the perspective of a static SoR and a revolving camera, a set of camera centers emerges that all produce the same contour generator. Such an equivalent set of camera centers is illustrated in Figure 4.7.

To describe the equivalence class of projection centers, we define the following equivalence relation:

$$(\mathbf{t}_a \sim \mathbf{t}_b) \leftrightarrow (\mathbf{t}_a = \mathbf{R}_z(\phi)\mathbf{t}_b \text{ for some angle } \phi).$$

It can be easily verified that this is a proper equivalence relation as it is reflexive ($\phi_a^a = 0$), symmetric ($\phi_b^a = -\phi_a^b$), and transitive ($\phi_c^a = \phi_b^a + \phi_c^b$). We define the *canonical representative* of an equivalence class for any class representative $t_o = (x, y, z)$ as

$$(\hat{d}, 0, \hat{h})^\top = \left(\sqrt{x^2 + y^2}, 0, z \right)^\top. \quad (4.13)$$

The canonical representative illustrates a reduction in degrees of freedom and simplifies the treatment of an arbitrary projection center.

4.4.3. Contour generator from generatrix

Combining the general contour generator constraints (Appendix 4.1) with the SoR parallel and meridian constraints (Section 4.3), the points ${}^o\mathbf{x}$ of the contour generator at height h are specifically the points of parallel \mathcal{P}_h with tangent planes that include the projection center ${}^o\mathbf{t}_c$. The coplanarity of points ${}^o\mathbf{t}_c$ and ${}^o\mathbf{x} \in \mathcal{P}_h$ on the tangent plane of ${}^o\mathbf{x}$ with normal ${}^o\mathbf{n}$ is described by the following system of equations:

$$\begin{aligned} {}^o\mathbf{n}^\top({}^o\mathbf{x} - {}^o\mathbf{t}_c) &= 0 \\ x^2 + y^2 &= r(h)^2. \end{aligned}$$

Solving this system of equations for x and y yields:

$$\begin{aligned} x &= (r^2 + rp(\hat{h} - h))/\hat{d} \\ y &= \pm\sqrt{r^2 - x^2}, \end{aligned} \tag{4.14}$$

with $r = r(h), p = \rho(h)$; see Appendix A.1.1 for full derivation.

This formulation admits two contour generator points ${}^o\mathbf{x}^+ = (x, +y, h)$ and ${}^o\mathbf{x}^- = (x, -y, h)$ due to the signed square root in Equation 4.14. These two solutions, the “left” and “right” contour points, are indeed expected and demonstrate that the contour generator has symmetry through the xz -plane. These forms are derived similarly to the rendering equations developed in Glachet et al. (1992).

4.4.4. Contour generator from occluding contour

Given image point \mathbf{x}_c , with corresponding homogeneous tangent line \mathbf{l}_c expressed on the canonical image plane with known projection center, their back-projection is constrained to

uniquely determine the corresponding generator point.

Transforming these homogeneous vectors into the SoR coordinate system, we have the direction vector

$${}^o\mathbf{x} = (x_o, y_o, z_o)^\top \quad (4.15)$$

of the camera ray

$${}^o\tilde{\mathbf{x}} = \lambda {}^o\mathbf{x} + {}^o\mathbf{t}_c \quad (4.16)$$

that must be tangent to and intersect the SoR surface and some point ${}^o\tilde{\mathbf{x}}$, and the homogeneous tangent line

$${}^o\mathbf{l} = (u'_o, v_o, w_o)^\top \quad (4.17)$$

that encodes the surface normal. The 3D normal component u_o along the x -axis is lost under projective division, yielding the projection u'_o . The normal is recovered as

$${}^o\mathbf{n} \propto \begin{pmatrix} u_o \\ v_o \\ w_o \end{pmatrix}, u_o = -\frac{y_ov_o + z_ow_o}{x_o} \quad (4.18)$$

using the tangent constraint Equation 4.2 (see Appendix A.1.2). The depth value λ of Equation 4.16 is recovered as

$$\lambda = \mu \hat{d}, \quad \mu = \frac{v_o}{u_oy_o - v_ox_o} \quad (4.19)$$

by applying the meridian constraint of Equation 4.10 (see Appendix A.1.3). With a known distance λ , the 3D contour generator point ${}^o\tilde{\mathbf{x}}$ is recovered by direct substitution. These equations are derived similarly to the reconstruction equations of Wong et al. (2002).

Using Equation 4.19 to introduce data term μ , the contour generator camera ray Equation 4.16 is rewritten as an explicit function of the representative camera center ${}^o\mathbf{t}_c = (\hat{d}, 0, \hat{h})$

of the form

$$\begin{pmatrix} \tilde{x}_o \\ \tilde{y}_o \\ \tilde{z}_o \end{pmatrix} = \mu \hat{d} \begin{pmatrix} x_o \\ y_o \\ z_o \end{pmatrix} + \begin{pmatrix} \hat{d} \\ 0 \\ \hat{h} \end{pmatrix} = \begin{pmatrix} (\mu x_o + 1) \hat{d} \\ \mu y_o \hat{d} \\ \mu z_o \hat{d} + \hat{h} \end{pmatrix} \quad (4.20)$$

4.4.5. Generatrix from contour generator

Given the 3D contour generator point ${}^o\tilde{\mathbf{x}}$ and surface normal ${}^o\mathbf{n}$ expressed in object coordinates, the generatrix values of $r(h)$ and $\rho(h)$ are uniquely determined. The generatrix height h is simply z_o , with

$$r(h) = \sqrt{x_o^2 + y_o^2}. \quad (4.21)$$

Equation 4.4 relates generatrix derivative ρ to 3D surface normal ${}^o\mathbf{n}$, allowing it to be recovered by normalizing w_o as

$$\rho(h) = \frac{w_o}{\sqrt{u_o^2 + v_o^2}}. \quad (4.22)$$

4.4.6. Generatrix and occluding contour bijection

Assuming a calibrated camera with known pose, a bijection between the generatrix and the occluding contour is formed by composition. To obtain the occluding contour from a known generatrix:

1. Compute 3D contour generator points from the generatrix using Section 4.4.3
2. Transform 3D contour generator points from object to camera coordinates
3. Project 3D contour generator points to the image plane

To recover the generatrix from an occluding contour:

1. Transform homogeneous image points and tangents to the canonical image plane
2. Compute 3D contour generator points using Section 4.4.4
3. Compute the generatrix from the 3D contour generator using Section 4.4.5

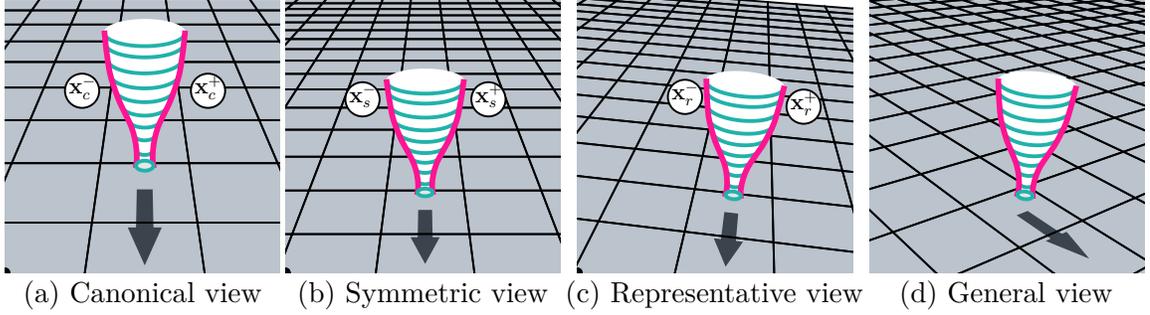


Figure 4.8: Forward-projection decomposition. The forward projection of an SoR is composed of a sequence of rotations. Pose recovery proceeds backwards from right to left.

4.5. Forward projection and pose decomposition

Recall from Equation 3.1 that object points ${}^o\mathbf{x}$ are expressed in camera a coordinates as

$${}^a\mathbf{x} = {}^a\mathbf{R}_o {}^o\mathbf{x} + {}^a\mathbf{t}_o,$$

where ${}^a\mathbf{R}_o$ and ${}^a\mathbf{t}_o$ represent the rotation and translation components of the pose, respectively. In calibrated camera coordinates, this equation also represents the forward projection of the object point into camera space. In prior work, the SoR pose in camera space is recovered via a myriad of means, with each method implicitly or explicitly parametrizing this equation in terms of the geometry it exploits.

For example, Glachet et al. (1992) uses bi-tangent points and cross sections to estimate the central-axis orientation $\hat{\mathbf{z}}$, restricting one of the three vectors of ${}^a\mathbf{R}_o$. Wong et al. (2002); Utcke and Zisserman (2003) partially recover ${}^a\mathbf{R}_o$ as a homography that is composed with a 1-parameter projective transform, representing an unknown SoR attitude. Wong et al. (2002) recovers this attitude using a visible cross section, and recovers ${}^a\mathbf{t}_o$ as a one degree-of-freedom translation vector using the rotational symmetry property of SoRs.

By assuming a known camera calibration matrix \mathbf{K} and examining all prior methods in Euclidean space, the various pose parametrizations can be seen to describe a common set pose variables. Explicitly decomposing the SoR pose in terms of these variables, the SoR

forward projection equation is factorized into five distinct transforms that each encode various geometric parameters:

$${}^g\tilde{\mathbf{x}} = \underbrace{{}^g\mathbf{R}_r}_1 \underbrace{{}^r\mathbf{R}_s}_2 \underbrace{{}^s\mathbf{R}_c}_3 \underbrace{{}^c\mathbf{R}_o}_4 \underbrace{{}^o\tilde{\mathbf{x}}}_5 + \underbrace{{}^g\mathbf{R}_r}^r \mathbf{t}_o. \quad (4.23)$$

Each of these factors (1-5) relate to previously discussed concepts that are used in some fashion in prior work.

1. The general pose is related to the representative pose by arbitrary rotation θ about $\hat{\mathbf{z}}$, yielding the 2-DoF representative center of projection ${}^r\mathbf{t}_o = (\hat{d}, 0, \hat{h})^\top$ (Section 4.4.2).
2. The representative pose is (non-uniquely) related to the symmetric pose by ${}^s\mathbf{R}_r$, encoding the 2-DoF camera roll and yaw, with arbitrary pitch.
3. The symmetric pose is related to the canonical pose by a camera pitch attitude of ϕ .
4. The canonical camera pose is related to the canonical SoR pose by a fixed transform (Section 4.4.1).
5. The contour generator is computed from SoR generatrix as a function of \hat{d} and \hat{h} (Section 4.4.3).

Transforms 1-3 (illustrated in Figure 4.8) encode five recoverable degrees of freedom (roll, pitch, attitude, distance, height), and one ambiguous degree of freedom (rotation about $\hat{\mathbf{z}}$). While various components of this decomposition are used in many works, the forward projection equation has not been previously represented in an explicit and fully parameterized form. This form is explored in subsequent chapters to accomplish pose recovery under a variety of novel conditions. Starting with a generic pose, each step of the forward projection when applied in reverse transforms the input to a more specialized pose, passing through the representative, symmetric, and canonical poses, in order.

4.5.1. Representative pose

A *general pose* is an arbitrary full 6-DoF 3D pose. It is reduced to a 5-DoF pose by considering only poses with a projection center in the form $(\hat{d}, 0, \hat{h})^\top$ in object coordinates. This *representative pose* utilizes the projective center equivalence relation defined in Section 4.4.2. The representative pose (Figure 4.8c) is transformed to the general pose (Figure 4.8d) by an arbitrary rotation θ about the central axis $\hat{\mathbf{z}}$. The image of the contour generator is invariant to such a rotation. All prior methods implicitly use the representative pose.

4.5.2. Symmetric pose

A *symmetric pose* is any camera pose where the camera’s z -axis intersects the central axis $\hat{\mathbf{z}}$. A representative pose is related to the symmetric pose by the 2-DoF rotation matrix ${}^s\mathbf{R}_r$ that encodes the camera roll and yaw. As shown in Section 3.3, the images of two cameras with the same camera center are related by a homography equal to the rotation relating them. In other words, an image of the SoR in the representative pose is transformed to the image of an SoR in the symmetric pose by the homography ${}^s\mathbf{R}_r$. The image of an SoR in the symmetric pose has symmetry through the projection of $\hat{\mathbf{z}}$ (Figure 4.8b), while the corresponding contour generator is symmetric through the xz -plane of the SoR coordinate system. Projectively transforming the input image to the symmetric pose, via rotation or homography, is the first step of all reconstruction techniques (Wong et al., 2002; Utcke and Zisserman, 2003; Colombo et al., 2005).

4.5.3. Canonical pose

The rotation of the canonical pose is a fixed transform relating the camera and the object coordinate systems, as defined in Section 4.4.1. The symmetric pose is related to the canonical pose by ${}^c\mathbf{R}_s$, a 1-DoF rotation of angle ϕ about the camera’s x -axis. By this relationship, the canonical pose is a specialization of a symmetric pose, where the camera’s z -axis both intersects and is perpendicular to the central axis $\hat{\mathbf{z}}$. The image of an SoR in the symmetric pose can be a linear stretching of the canonical image (Figure 4.8b).

The rotation between the representative image and the canonical image of an SoR is by the definition of the bijection between the occluding contour, contour generator and SoR generatrix. The transform from the symmetric pose to the canonical pose is computed implicitly for any approach that computes the SoR attitude (Glachet et al., 1992; Wong et al., 2002; Colombo et al., 2005).

4.6. Symmetric axis recovery via cylindrical projection

A point on the infinite unit cylinder with central axis $\hat{\mathbf{z}}$ is expressed in cylindrical coordinates as $(\theta, h)^\top$, where θ is the azimuth angle and h is the height along $\hat{\mathbf{z}}$. A point ${}^o\mathbf{x} = (x, y, z)^\top$ on the unit image sphere is projected to the cylinder as

$$(\theta, h)^\top = (\arctan2(y, x), z)^\top \quad (4.24)$$

with an inverse projection of

$$(x, y, z)^\top = (\cos \theta, \sin \theta, h). \quad (4.25)$$

The representation of an SoR under this axis-aligned cylindrical projection exhibits several useful properties. Given that the central axis of the SoR is parallel with the central axis of the cylinder, the following statements are true:

1. The SoR central axis $\hat{\mathbf{z}}$ projects to a vertical line at azimuth θ on the cylinder.
2. The SoR's occluding contours are symmetric through the vertical line at azimuth θ .

Assuming the normal \mathbf{n}_z of $\hat{\mathbf{z}}$ is known, then the rotation $\mathbf{R}_{\hat{\mathbf{z}}}$ that aligns the SoR's axis to a canonical upright cylinder is represented in angle-axis form with angle

$$\theta = \arccos \left(\frac{\mathbf{n}_z \cdot (0, 0, 1)}{\|\mathbf{n}_z\|} \right)$$

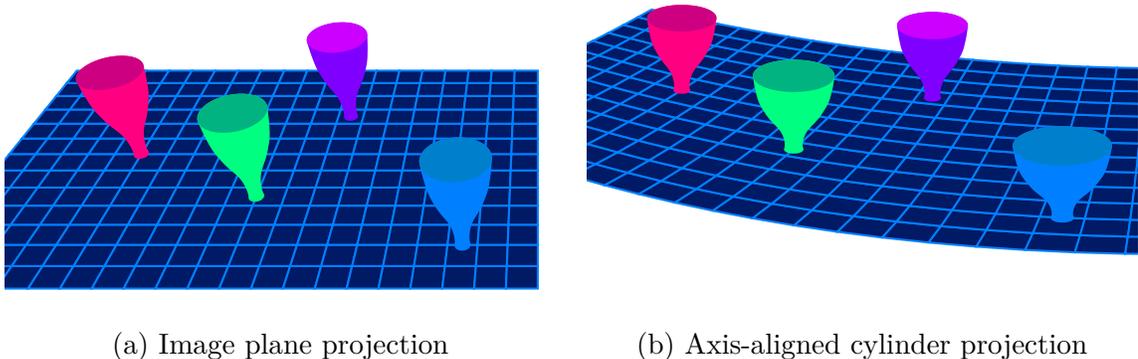


Figure 4.9: Symmetry of an SoR under cylindrical reprojection. SoRs with the same central axis normal appear to have different tilts on the image plane (a) do to perspective distortion. However, when they are projected onto a cylinder that is aligned with the central axis, they all are represented as symmetric, with a height that is proportional to the radial distance from the camera center.

and axis

$$\mathbf{v} = \mathbf{n}_z \times (0, 0, 1).$$

After the image sphere is rotated by \mathbf{R}_z and projected to the cylinder, any SoR with $\hat{\mathbf{z}}$ parallel to \mathbf{n}_z will appear as symmetric, regardless of translation from the camera (see Figure 4.9). As SoRs move away from the camera, they will appear to translate upwards in the cylindrical image. The cylindrical image provides an additional convenience in that it can be unwrapped and represented as a flat image, allowing for traditional image processing techniques.

To locate the image of the SoR central axis, it is sufficient to examine a single cylinder height (horizontal scanline) and identify the azimuth coordinates of the two points that are symmetric with respect to the axis. The azimuth angle θ of the symmetric axis is simply the midpoint these two azimuth coordinates. If the SoR contours are known, then such midpoint azimuths are obtained by averaging points with corresponding cylinder heights. If the contours are not known, then a symmetry operator such as defined in Reisfeld et al. (1995) can be applied to produce azimuth angle hypotheses. Any such azimuth θ back-

projects to a line l in the image plane.

Part III

Models and Methods

5 Datasets

This chapter presents three datasets that are designed to evaluate the performance of the pose and shape recovery techniques of the major thesis contributions. The effect of several variables on performance is measured with respect to the accuracy of the recovered pose parameters (\hat{d}, \hat{h}, ϕ) , and shape (generatrix radii) as compared to the groundtruth.

Specifically, the three datasets;

- View-plane dataset – images of a single opaque SoR taken within a plane of camera positions (\hat{d}, \hat{h}) ,
- Synthetic view-plane dataset – synthetic images in the same view-plane produced with various parameters,
- Transparent SoR dataset – images of a single transparent SoR taken in various scenes and poses,

are used to evaluate the effect on performance of the variables;

- Noise – sensor noise, contour discretization and localization,
- SoR shape – generatrix radii $r(h)$ of heights h ,
- Camera position – camera height \hat{h} and distance \hat{d} from the SoR origin,
- Camera baseline – difference in camera height $\Delta\hat{h}$ and depth $\Delta\hat{d}$ between two views,
- Number of views – image sample count k from which evidence is aggregated.

5.1. View-plane dataset

Camera pose is an important parameter for experimental evaluation. In the single-view context, the most relevant aspects of the camera pose is the camera height \hat{h} and distance

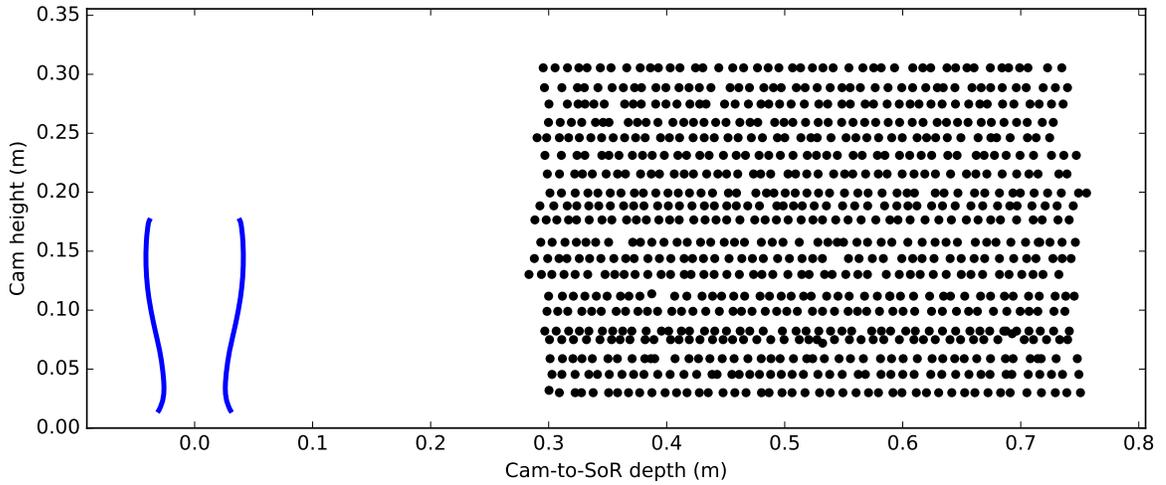
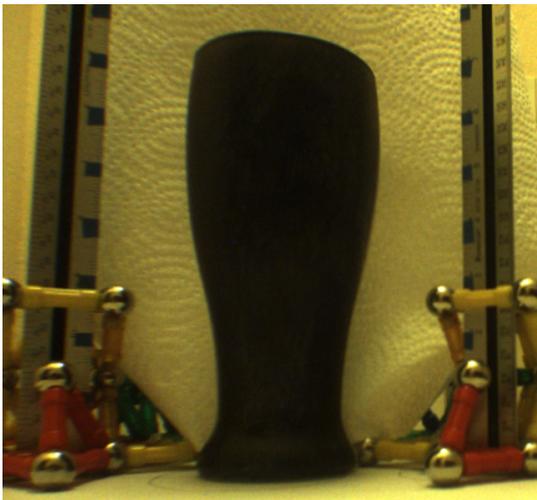
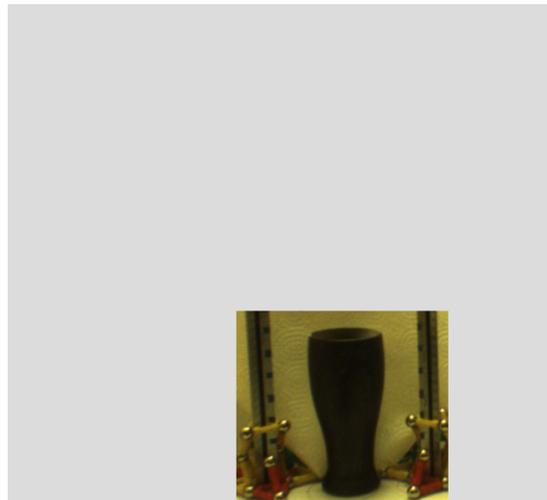


Figure 5.1: View-plane sample points with respect to SoR origin. The representative camera pose is parameterized as a function of camera depth \hat{d} from the SoR origin and height \hat{h} above it. The view-plane dataset samples this space with images taken from positions (\hat{d}, \hat{h}) that approximately form a grid. These sample points are visualized as black dots, shown in relation to the SoR with blue contours.



(a) Near-view sample



(b) Far-view sample

Figure 5.2: Extremal views from the sampled view plane. Image (a) shows the nearest view with respect to the SoR origin, while image (b) shows the farthest. Represented with the same scale, these images show the variability in SoR apparent size over the sampled view plane. The vertical rulers are perpendicular to the supporting plane and in a known configuration with respect to the SoR. They allow for accurate normal, scale, and translation estimation of the SoR pose.

\hat{d} from the SoR origin. In a multi-view context, the baseline between views is also an important attribute, defined as the difference in depth $\Delta\hat{d}$ and height $\Delta\hat{h}$.

To measure the effects of these parameters, a dense sampling of the parameter space is captured with a calibrated Point Grey Bumblebee color camera for a single SoR. Such sample points, of the form (\hat{d}, \hat{h}) , approximate a grid within a plane of image views along the depth and height axes (see Figure 5.1), called a *view plane*. There are 900 view-plane samples with 45 depth values in 100mm increments and 20 height values in 135mm increments spanning a 45cm×27cm area. This range provides good variability in the apparent size of the SoR, with the closest view to the SoR (Figure 5.2a) at a depth of 30cm and a height of 3cm, and the farthest view (Figure 5.2b) at a depth of 85cm and a height of 30cm.

The groundtruth SoR poses are estimated using two vertical rulers that are perpendicular to the supporting plane and in a known configuration with respect to the SoR. Identification of the two parallel ruler lines in the image provide the 3D supporting plane normal as well one degree of freedom of the SoR’s origin. The ruler markings provide scale and height information to fully determine the translation between the camera and SoR.

Color threshold segmentation is used to compute the apparent contours for all views. Generatrix reconstructions are estimated from these contours for all views, and the groundtruth generatrix is taken as the mean generatrix over all reconstructions. Figure 5.3a shows all 900 estimated generatrices along with the mean reconstruction and standard deviation bounds. Examining the generatrix closely in Figure 5.3b and Figure 5.3c, most generatrices fall within the 0.4mm standard deviation from the mean generatrix. This consistency lends confidence to the groundtruth pose estimations and also provides a baseline for the expected variance from the groundtruth for any reconstruction experiments.

5.2. Synthetic view-plane dataset

The view-plane dataset allows for the evaluation of camera pose parameters on actual sensor data, however it fails to provide a means to tractably test several other parameters

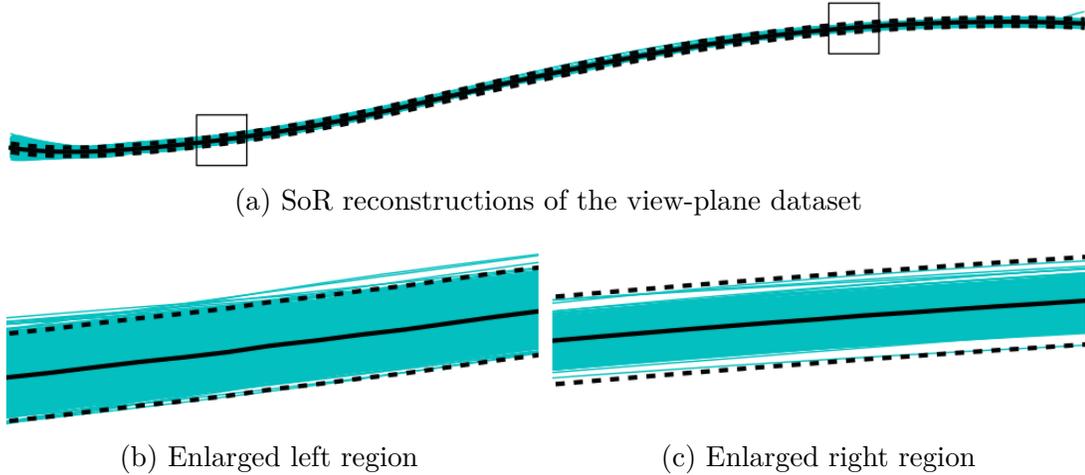


Figure 5.3: Reconstructed SoR generatrices from the view-plane dataset with aggregated statistics. All reconstructed generatrices from the view-plane dataset are presented in (a), with the solid black line representing the average reconstruction and the dotted black lines representing the computed standard deviation of 0.4mm from this average reconstruction. Two regions on the left and right side of (a) are outlined with a black box and enlarged in (b) and (c) for greater detail. Notice that most reconstructions fall within the black dotted lines representing standard deviation in radii over all reconstructions.

of interest. To conduct experiments over varying sensor noise models and SoR shapes, the view-plane dataset is replicated synthetically for every desirable test configuration.

Zero-mean Gaussian noise parameterized by variance σ is combined with optional synthetic apparent contour discretization to yield 10 different sensor noise models, as enumerated in Table 5.1. A database of 49 SoR generatrices created from real objects (see Figure 5.4) produces a large experimental parameter space, with $(900 \text{ poses}) \times (10 \text{ noise models}) \times (49 \text{ shapes})$ choices yielding a half-million single-view configurations. This number is on the order of hundreds of millions when considering two-view configurations. Due to this multiplicity, at least one of these parameters is typically held constant in experimental setups while the others are explored.



Figure 5.4: Synthetic SoR models and corresponding identification numbers. The SoRs corresponding to 49 generatrices that compose the synthetic dataset are represented in this figure, shown above their corresponding identification numbers. SoR-34 represents the object that is imaged in the real dataset views.

Noise model	Discretized?	Sensor σ (px)
Synth-0	No	0.00
Synth-1	Yes	0.00
Synth-2	Yes	0.25
Synth-3	Yes	0.50
Synth-4	Yes	0.75
Synth-5	Yes	1.00
Synth-6	Yes	1.25
Synth-7	Yes	1.50
Synth-8	Yes	1.75
Synth-9	Yes	2.00

Table 5.1: Synthetic contour noise model parameterizations. Synthetic views are generated with one of ten noise models, parameterized by the variance σ of a zero-mean Gaussian and whether the output signal is continuous or discretized. Such discretization simulates the effect of sensor pixel sampling.

5.3. Transparent stereo dataset



Figure 5.5: Transparent SoR stereo setup exemplar images. The transparent stereo dataset is composed of one SoR seen from multiple viewpoints in six scenes and two configurations, both empty (top row) and filled with water (bottom row). The corresponding scene identifier is presented underneath each column.

A stereo dataset of transparent objects is used to evaluate the performance of a stereoscopic cue for the reconstruction of optically challenging surfaces of revolution. This dataset contains 120 images containing a single SoR (SoR-34 as identified in the synthetic dataset) as seen from 10 known camera poses in 12 different scene configurations. These configurations are shown in Figure 5.5 and include both an empty and water-filled glass placed within six different background settings.

Four settings have background clutter objects resting atop a black or wood textured supporting surface. The remaining two settings are clutter-free, one with a plain black surface and the other with a speckled surface texture. The pose of the SoR and supporting plane is known via the use of a checkerboard calibration target that is imaged with a static camera and swapped out for the object in a known relative position.

6

Symmetry rotation recovery and 3D axis triangulation

The identification of the SoR central-axis image projection provides two degrees of freedom of the SoR pose, called the symmetric rotation. Given two SoR views from cameras that are related by a known transformation, the 3D SoR central-axis is triangulated by back-projecting the 2D SoR central-axes. This chapter presents and evaluates the techniques for accomplishing these two tasks.

6.1. Two-point minimal problem for symmetry recovery

The pose of an SoR is decomposed in Section 4.5 to yield two parametrizable transforms. The first is ${}^s\mathbf{R}_r$, which transforms the SoR's xz -plane to be coplanar with the camera's yz -plane by encoding the camera's roll and yaw. The second is ${}^c\mathbf{R}_s$, encoding the camera's attitude ϕ and aligning the camera's z -axis to be perpendicular to the SoR's z -axis.

From the definition in Section 4.4.3, the contour generators of an SoR are symmetric through the SoR's xz -plane. From the bijection of Section 3.3, it is equivalent to consider occluding contours as the projection of the contour generator onto the image sphere. Under such a spherical projection, the 3D contour generators that are symmetric through the xz -plane project to 2D occluding contours that are symmetric through the great circle representing the projection of the SoR's central axis, $\hat{\mathbf{z}}$.

Exploiting this property of occluding contour symmetry, the xz -plane normal, the projection of $\hat{\mathbf{z}}$, and associated symmetric rotation are recovered using a two-point minimal problem (Phillips and Daniilidis, 2016). If two occluding contour points \mathbf{a} and \mathbf{b} are known to be

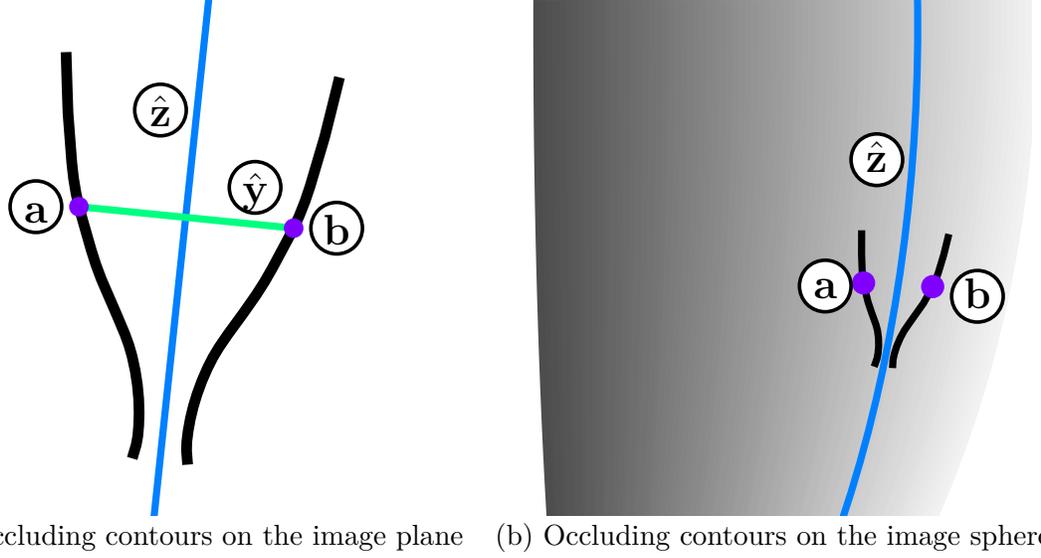


Figure 6.1: Symmetric points and axes for the symmetry rotation minimal problem. Two corresponding SoR points a and b are homologous on the image plane (a) through the projection of central axis $\hat{\mathbf{z}}$, and are symmetric through the great circle projection of $\hat{\mathbf{z}}$ onto the image sphere (b). The axis $\hat{\mathbf{y}}$ represents the normal of the great circle. From two points alone, it is not possible to find the attitude of $\hat{\mathbf{z}}$, which may point towards anywhere on the great circle.

symmetric correspondences, then the pose of $\hat{\mathbf{z}}$ is recovered up to an unknown attitude, providing transform ${}^s\mathbf{R}_r$.

6.1.1. Two-point minimal problem

Given two SoR occluding contour points \mathbf{a} and \mathbf{b} that are homologous on the image plane (Figure 6.1a) and therefore symmetric on the image sphere (Figure 6.1b), the symmetric rotation is recovered as follows.

The y -axis is the normal of the xz -plane and its corresponding great circle on the image sphere. It must be parallel to the line segment connecting \mathbf{a} and \mathbf{b} , so it is recovered as

$$\hat{\mathbf{y}} = \frac{\mathbf{b} - \mathbf{a}}{\|\mathbf{b} - \mathbf{a}\|}. \quad (6.1)$$

The z -axis is projection of the central axis $\hat{\mathbf{z}}$, with unknown attitude ϕ relating it to the

true central axis. It must be orthogonal to \mathbf{a} and \mathbf{b} on the image sphere, respecting

$$\hat{\mathbf{z}} = \frac{\mathbf{a} \times \mathbf{b}}{\|\mathbf{a} \times \mathbf{b}\|}. \quad (6.2)$$

To complete the orthonormal basis ${}^s\mathbf{R}_r = (\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})^\top$, the remaining x -axis is constrained to be

$$\hat{\mathbf{x}} = \frac{\hat{\mathbf{y}} \times \hat{\mathbf{z}}}{\|\hat{\mathbf{y}} \times \hat{\mathbf{z}}\|}. \quad (6.3)$$

The symmetry rotation is explicitly defined in terms of corresponding points \mathbf{a} and \mathbf{b} as

$${}^s\mathbf{R}_r = \left(\begin{array}{c} \hat{\mathbf{y}} \times \hat{\mathbf{z}} \\ \frac{\hat{\mathbf{y}} \times \hat{\mathbf{z}}}{\|\hat{\mathbf{y}} \times \hat{\mathbf{z}}\|}, \underbrace{\frac{\mathbf{b} - \mathbf{a}}{\|\mathbf{b} - \mathbf{a}\|}}_{\hat{\mathbf{y}}}, \underbrace{\frac{\mathbf{a} \times \mathbf{b}}{\|\mathbf{a} \times \mathbf{b}\|}}_{\hat{\mathbf{z}}} \end{array} \right)^\top. \quad (6.4)$$

The reflection homology, \mathbf{H}_z , that maps one occluding contour onto the other is defined using ${}^s\mathbf{R}_r$ as

$$\mathbf{H}_s = {}^s\mathbf{R}_r^\top \mathbf{D} {}^s\mathbf{R}_r. \quad (6.5)$$

Considering a point being transformed from right to left, ${}^s\mathbf{R}_r$ rotates the point to be symmetric about the camera's y -axis, the reflection matrix

$$\mathbf{D} = \text{diag}(1, -1, 1) \quad (6.6)$$

reflects it about the y -axis, and ${}^r\mathbf{R}_s$ rotates this reflected point back to the representative pose.

6.1.2. Minimal problems and parameter estimation

Minimal problems allow for the estimation of the parameters encoded by a potentially large set of data points using a minimal number of data points. In the case of the symmetric estimation problem, only two symmetric points are required to recover the axis of symmetry

that is respected by the symmetric apparent contours. If the minimal problem is applied to two apparent contour points that are not symmetric, then the resulting parameters will be incorrect. Minimal problems therefore require “correct” input points that encode the parameters being estimated in order to produce a valid solution. Without a priori knowledge of such points, one strategy to parameter estimation follows the insight from Fischler and Bolles (1981) and the now classic RANSAC algorithm. The procedure for an n -point minimal problem is outlined as follows:

1. Sample an n -point input from some subset \mathcal{S}_{input} of the input point power set
2. Use the minimal problem to compute the corresponding parameter model Ω
3. Evaluate the parameter fit over all data points using some error function f_{err}
4. Repeat steps 1-3 some stop criterion is reached
5. Use the parameter model $\hat{\Omega}$ with the best fit over all data points
6. Optionally refine the parameter model over all data points using f_{ref} .

The choice of input space \mathcal{S}_{input} , error function f_{err} , stopping criteria, and f_{ref} are implementation details that are specific to the problem domain. For problems in which two contours \mathbf{A} and \mathbf{B} are aligned by the application of model parameters Ω , it is useful to define a nearest-neighbor correspondence set \mathcal{C}_{nn} . This set contains the indices (i, j) of contour points $a_i \in \mathbf{A}$ and $b_j \in \mathbf{B}$ that are less than d_{max} apart after transformed by Ω . If $|\mathcal{C}|$ is the number of correspondences, then the error function may be expressed as

$$f_{err} = \begin{cases} \infty & |\mathcal{C}_{nn}| < k \\ \frac{1}{n} \sum_i^n |\mathbf{a}_i - \Omega(\mathbf{b}_i)|_2 & \text{otherwise} \end{cases},$$

where k is the minimal required number of correspondences and $\Omega\mathbf{b}_i$ is the point \mathbf{b}_i remapped by parameter model Ω to align with \mathbf{a}_i .

6.1.3. Symmetric rotation estimation and refinement

Given two contours \mathbf{A} and \mathbf{B} that are assumed to belong to the profile of an SoR, the two-point minimal problem provides a means for generating hypotheses for ${}^s\mathbf{R}_r$ by applying

Equation 6.4 to any point pair $(\mathbf{a}_*, \mathbf{b}_*) \in \{\mathbf{A} \times \mathbf{B}\}$. The homology \mathbf{H}_z corresponding to rotation ${}^s\mathbf{R}_r$ maps a point \mathbf{b} onto \mathbf{a} as $\mathbf{H}_z\mathbf{b}$. Using this mapping to specialize the correspondence residual function of the hypothesis search procedure in Section 6.1.2 as

$$\frac{1}{n} \sum_i^n |\mathbf{a}_i - \mathbf{H}_z \mathbf{b}_i|_2, \quad (6.7)$$

the best unrefined rotation estimate is recovered and denoted as $[{}^s\mathbf{R}_r]_0$.

This initial rotation can be iteratively refined following the assumption that the rotation estimate $[{}^s\mathbf{R}_r]_k$ yields corresponding contours ${}^k\mathbf{A}$ and ${}^k\mathbf{B}$ that are closely aligned, yet related by a small refinement rotation \mathbf{R}_Δ as

$${}^k\mathbf{a} = \mathbf{R}_\Delta {}^k\mathbf{b}, \quad (6.8)$$

which decreases the correspondence residual

$$\frac{1}{n} \sum_i^n |{}^k\mathbf{a}_i - \mathbf{R}_\Delta {}^k\mathbf{b}_i|_2. \quad (6.9)$$

By repeated estimation and application of such refinement matrices, the symmetric rotation is iteratively improved as

$$[{}^s\mathbf{R}_r]_{k+1} = [\mathbf{R}_\Delta]_k [{}^s\mathbf{R}_r]_k. \quad (6.10)$$

Symmetric rotation refinement is summarized as follows:

1. Use the current symmetric rotation estimate $[{}^s\mathbf{R}_r]_k$ to compute aligned nearest correspondence contours ${}^k\mathbf{A}$ and ${}^k\mathbf{B}$.
2. Compute $[\mathbf{R}_\Delta]_k$ that minimizes the correspondence residual (Equation 6.9).
3. Rotate $[{}^s\mathbf{R}_r]_k$ by \mathbf{R}_Δ to yield the new estimate $[{}^s\mathbf{R}_r]_{k+1}$.
4. Repeat steps 1-3 until convergence or iteration stop criteria is reached.

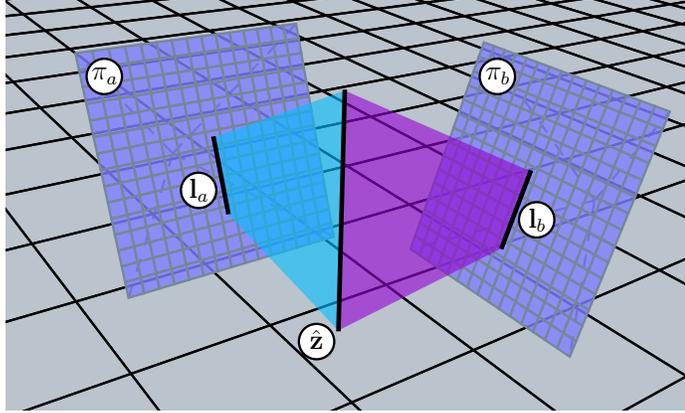


Figure 6.2: Central axis triangulation by intersection of back-projection planes. Since the relative pose of the two image planes π_a and π_b are known, the spatial relationship between the measured image lines \mathbf{l}_a and \mathbf{l}_b are also known. These lines back-project into planes that intersect at the 3D line $\hat{\mathbf{z}}$.

See Appendix B.1 for the computation of \mathbf{R}_Δ .

6.2. Two-view 3D central axis triangulation

Given two calibrated stereo views with known 2D central-axis projections, the 3D central-axis is recovered from the intersection of two back-projected planes. Section 6.1 provides a means for recovering the 2D projection of the SoR central axis $\hat{\mathbf{z}}$. Recall from Section 3.2.5 that the back-projection of a 2D line is a 3D plane through the camera center, the normal \mathbf{n} of which is equal to the homogeneous line representation \mathbf{l} in the calibrated camera coordinate system. A calibrated stereo camera system provides the relative point transform between camera a and camera b as

$${}^a\mathbf{x} = {}^a\mathbf{R}_b {}^b\mathbf{x} + {}^a\mathbf{t}_b \quad (6.11)$$

and the relative normal transform as

$${}^a\mathbf{n} = {}^a\mathbf{R}_b {}^b\mathbf{n}, \quad (6.12)$$

where ${}^a\mathbf{R}_b$ is the rotation from camera b to camera a , and ${}^a\mathbf{t}_b$ is the origin of camera b expressed in the coordinate system of camera a .

Using these transforms, the 3D central axis is triangulated by computing the intersection of the two back-projection planes in the same coordinate system, as shown in Figure 6.2. The normals of the two planes expressed in camera a are expressed as

$${}^a\mathbf{n}_a = {}^a\mathbf{l}_a \quad (6.13)$$

$${}^a\mathbf{n}_b = {}^a\mathbf{R}_b {}^b\mathbf{l}_b, \quad (6.14)$$

where the left superscript denotes coordinate frame and the subscript denotes the entity. The planes must contain the camera centers, yielding two known plane points

$${}^a\mathbf{p}_a = \mathbf{0} \quad (6.15)$$

$${}^a\mathbf{p}_b = {}^a\mathbf{t}_b. \quad (6.16)$$

The line of intersection of these two plane-point representations yields the $\hat{\mathbf{z}}$ direction vector \mathbf{n}_z as well as plane point \mathbf{p}_z . The attitude ϕ of ${}^c\mathbf{R}_s$ that maps the symmetric pose to the canonical pose is directly recovered from \mathbf{n}_z as the signed angle between $(0, 1, 0)^\top$ and $\mathbf{R}_s^\top \mathbf{n}_z$.

The representative center of projection ${}^r\mathbf{t}_o = (\hat{d}, 0, \hat{h})$ is directly recovered from \mathbf{p}_z via Equation 4.13, where \hat{h} specifies an arbitrary reference point for height in the case that there is no generatrix known a priori. After triangulation, all missing pose parameters of the forward projection Equation 4.23 are specified, yielding the SoR pose ${}^c\mathbf{F}_o$ in the camera frame and allowing for recovery of the SoR generatrix.

6.3. Surface of revolution generatrix recovery

The bijections derived in Section 4.4 provide a means of recovering the contour generator and generatrix $r(h)$ of an SoR from the occluding contour image points and tangents, assuming knowledge of its pose in the camera frame \mathbf{F}_c^o as well as camera intrinsics \mathbf{K} .

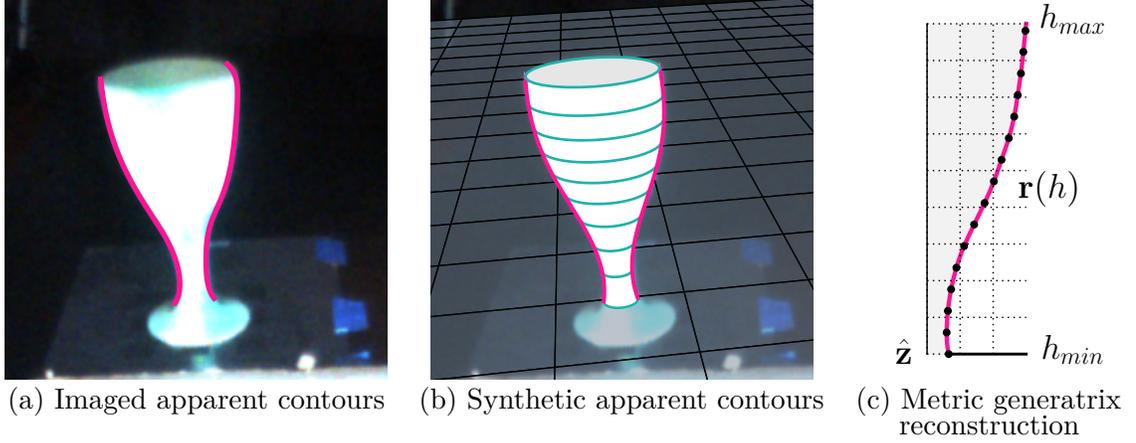


Figure 6.3: Metric generatrix reconstruction from occluding contour. Figure a. shows the apparent contours of an SoR with known pose from a camera with known camera intrinsics. The resulting generatrix from the Euclidean-based reconstruction is shown in figure c, which is used to generate the synthetic apparent contours of figure b.

In this chapter, the pose ${}^c\mathbf{F}_o$ is recovered piecewise via the forward-projection decomposition equation, 4.23, as follows:

1. Symmetric rotation ${}^s\mathbf{R}_r$: The 2D projection of the central axis $\hat{\mathbf{z}}$, and therefore the symmetric rotation is estimated using the two-point minimal problem presented in Section 6.1.
2. Camera attitude ϕ : The known relative camera transform ${}^a\mathbf{F}_b$ is used to triangulate the 3D-axis in Section 6.2 by intersecting the back-projection planes of the estimated central axis, yielding a line in point-vector form. The vector directly encodes ϕ .
3. Camera center $(\hat{d}, 0, \hat{h})^T$: Any point on the recovered 3D axis line yields \hat{d} when rotated into the canonical pose using ϕ to parameterize the rotation ${}^c\mathbf{R}_s$. The camera height \hat{h} is a free variable and can be arbitrarily chosen.

The requirement of a known relative camera transform to recover the SoR pose with respect to each image is relaxed in Chapter 8, with the above steps 2 and 3 replaced with minimal problems using occluding contour points and tangents.

Regardless of the method used for SoR pose recovery, generatrix recovery is summarized as follows:

1. Transform points and tangents to the calibrated coordinate system using \mathbf{K}^{-1} .
2. Compute the SoR canonical pose in the camera, ${}^c\mathbf{F}_r$ via whichever method available.
3. Transform points and tangents to the canonical view via homography ${}^c\mathbf{H}_r = {}^c\mathbf{R}_r$.
4. Compute the contour generator via the constraints of Section 4.4.4.
5. Compute the generatrix as $r(z) = \sqrt{x^2 + y^2}$ for each generator point $\tilde{\mathbf{x}}_o = (x, y, z)$.

Figure 6.3 shows an example recovered generatrix from a real image using this method.

6.4. Symmetry rotation recovery evaluation

The first transform that is estimated in the forward-projection decomposition is the 2-DoF symmetric rotation that encodes the projection of the SoR central axis in the image. The accuracy of the estimated symmetric rotation is expressed with respect to the groundtruth by two different metrics. The first metric is the average distance between the estimated and groundtruth central-axis projections in the image, expressed in pixels. This encompasses both error in axis orientation and translation. The second metric is the angle between these two central axis projections on the calibrated camera plane, capturing only orientation information.

The performance of the symmetric rotation estimation technique is evaluated with respect to sensor noise and camera position. This is achieved by the use of both real and synthetic datasets defined over a plane of camera view locations. Additionally, the improvement due to the iterative infinitesimal symmetric pose refinement technique is quantified for a fixed SoR over the view plane.

For two apparent contours \mathbf{A} and \mathbf{B} , the space of input points for the minimal problem is defined as the Cartesian product of all points in \mathbf{A} and every tenth point in \mathbf{B} .

Noise model	Distance error (px)		Angle error (deg)	
	Initial	Refined	Initial	Refined
Real	0.09	0.07	0.04	0.03
Synth-0	0.03	0.01	0.01	0.00
Synth-1	0.49	0.49	0.02	0.01
Synth-2	0.51	0.50	0.02	0.01
Synth-3	0.53	0.51	0.02	0.01
Synth-4	0.55	0.53	0.04	0.02
Synth-5	0.83	0.70	0.24	0.15
Synth-6	0.94	0.78	0.27	0.19
Synth-7	0.86	0.75	0.19	0.15
Synth-8	0.93	0.78	0.23	0.13
Synth-9	1.04	0.88	0.29	0.18

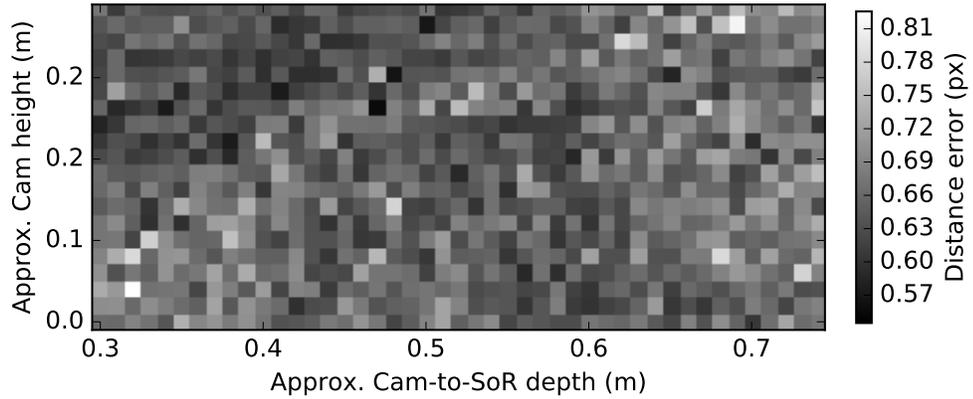
Table 6.1: Effect of iterative refinement on symmetric pose geometric error. The geometric error of a given symmetric axis is measured in terms of the average axis distance in pixels and the angular distance in degrees from the groundtruth axis of symmetry. Summarized in this table is the error of the initial estimate versus the iteratively refined estimate for the real and synthetic datasets over the range of noise models.

6.4.1. Effect of sensor noise model and iterative refinement

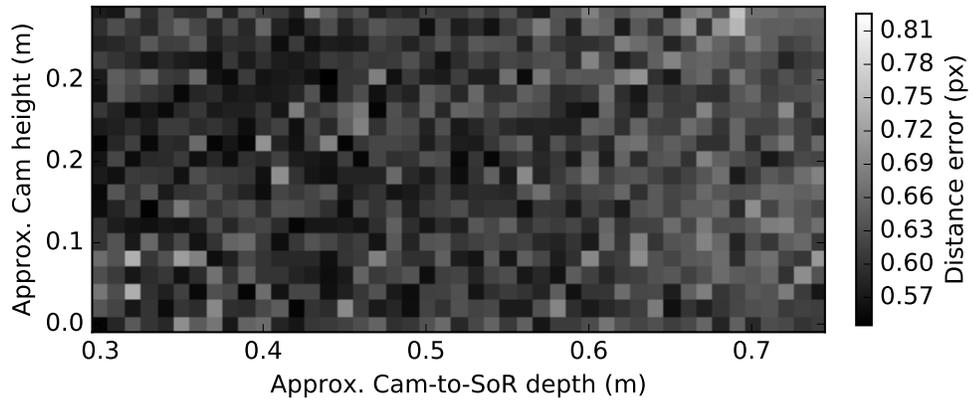
The synthetic view-plane dataset defines 10 noise models (see Figure 5.1). Combining the 900 SoR images of the view-plane dataset with the simulated images of the synthetic view-plane dataset, the effect of sensor noise on symmetric pose estimation is evaluated with 11 sensor models on 900 views each for a total of 9,900 sample points. Table 6.1 presents the mean pixel distance and axis-angle error for each sensor model, both before and after refinement by the iterative infinitesimal rotation. Performance on the real sensor data is similar to the Synth-4 model with respect to angle error. With respect to pixel distance error, it performs similarly to the Synth-0 and Synth-1 models. This suggests that the contour localization error on the real dataset is bounded by 0.75, the error used for the Synth-4 noise model. The improvement due to refinement becomes much more appreciable as the sensor noise increases, with a greater improvement in angular error over pixel distance error.

6.4.2. Effect of camera position and iterative refinement

Camera positions that are expressed in the representative pose span a plane with points parameterized by depth \hat{d} and height \hat{h} as measured with respect to the SoR origin. The

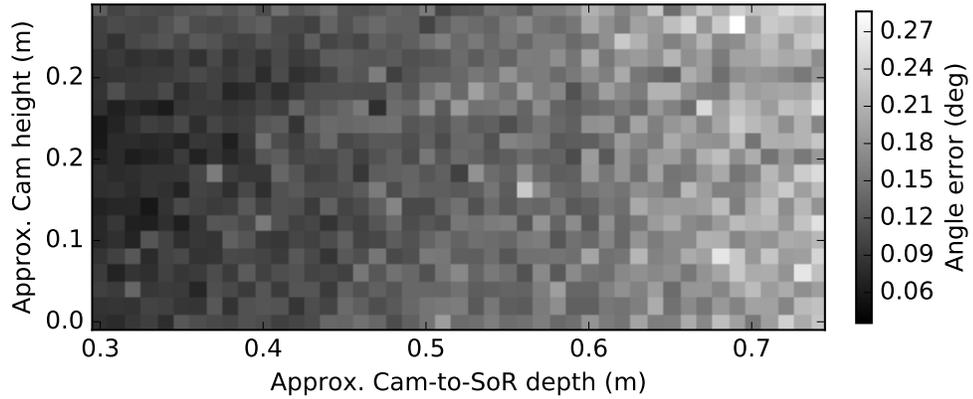


(a) Initial symmetric axis pixel errors

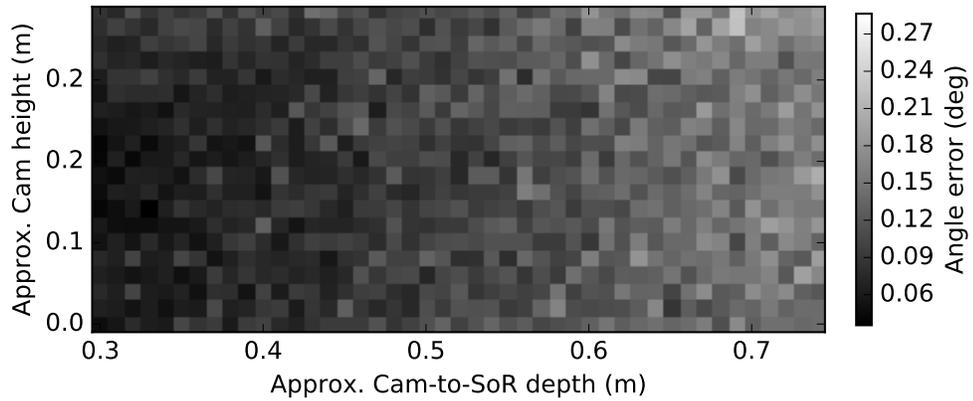


(b) Refined symmetric axis pixel errors

Figure 6.4: Estimated and refined symmetric axis distance errors per view-plane sample point. The error in symmetric axis angle (in pixels as indicated by the color bar) for the initial (a) and refined (b) poses are aggregated over all 49 SoR models with synthetic views generated using the Synth-4 noise model. Each square represents the mean of 49 error values for a camera with depth \hat{d} (x -axis) and \hat{h} (y -axis), with darker values corresponding to smaller errors.



(a) Initial symmetric axis-angle errors



(b) Refined symmetric axis-angle errors

Figure 6.5: Estimated and refined symmetric axis-angle errors per view-plane sample point. The error in symmetric axis angle (in degrees as indicated by the color bar) for the initial (a) and refined (b) poses are aggregated over all 49 SoR models with synthetic views generated using the Synth-4 noise model. Each square represents the mean of 49 error values for a camera with depth \hat{d} (x -axis) and \hat{h} (y -axis), with darker values corresponding to smaller errors.

synthetic view-plane dataset provides 900 views of 49 SoR models that are further parameterized by any one of 10 sensor noise models. The effect of camera position on symmetric pose estimation is evaluated over all 900 views and 49 SoR models with the Synth-4 noise model, yielding 44,100 sample points.

Figures 6.4 and 6.5 show the average pixel distance and angle error, respectively, for each of the 900 camera positions on the view plane. They show that estimation error tends to increase as the distance of the camera from the SoR increases. This relationship is explained by the fact that as the apparent size of the object decreases, the sensor noise model remains constant. The signal-to-noise ratio therefore decreases as an object recedes into the distance, resulting in the degradation of the symmetry estimation quality.

Figures 6.4b and 6.5b show the improvement due to iterative refinement over the initial estimates in Figures 6.4a and 6.5a. The average pixel distance and angle errors for the initial estimates are 0.65 pixels and 0.14° versus 0.62 pixels and 0.11° , respectively, for the refined estimate, showing a marginal improvement.

6.5. Two-view 3D axis triangulation evaluation

The pose of an SoR is triangulated from two views that are related by a known relative transform. Once the symmetric poses of the two views are estimated, the recovered 2D projections of the central axes are back-projected to two planes that intersect in a line containing the 3D SoR central axis. The accuracy of the 3D axis triangulation technique is highly dependent on the accuracy of the symmetric axis estimation.

The accuracy of the triangulated axis is measured with respect to the groundtruth by two different metrics, capturing error in translation and orientation. Since the position along the central axis is arbitrary, a 2-DoF translation error is computed that encodes the axis bearing and distance from the camera. The orientation error is computed as the angle between the estimated and groundtruth axis direction vectors. The performance of the axis triangulation technique is evaluated with respect to three factors: the sensor noise model,

Noise model	Translation error (mm)	Axis-angle error (deg)
Real	3.1	2.2
Synth-0	0.8	0.9
Synth-1	2.5	1.2
Synth-2	2.8	1.3
Synth-3	2.9	1.3
Synth-4	4.6	2.1
Synth-5	26.5	9.0
Synth-6	30.4	9.6
Synth-7	30.7	9.1
Synth-8	27.9	8.6
Synth-9	23.7	6.6

Table 6.2: Triangulated pose translation and axis estimation errors by dataset. The recovered SoR pose is described by its central axis direction and the 2-DoF translation encoding its bearing and distance from the camera. These values are compared with the groundtruth pose locations and aggregated over all 67081 pose configurations per dataset and presented as the mean error.

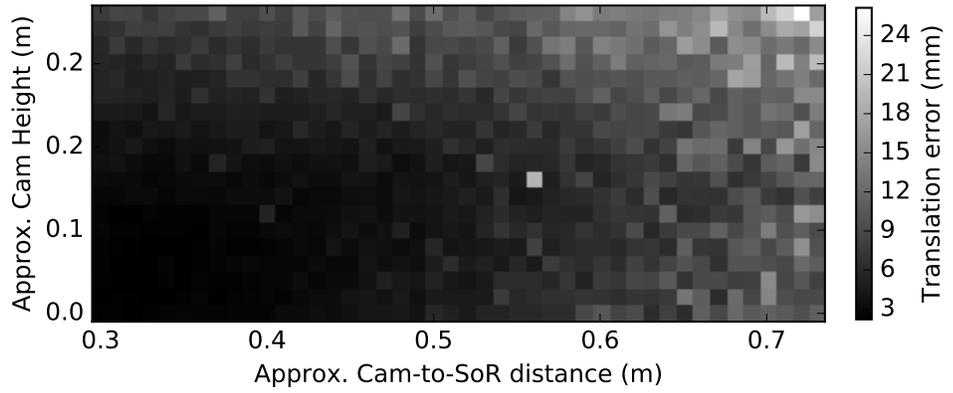
camera position, and camera pair baseline.

6.5.1. Effect of sensor noise model

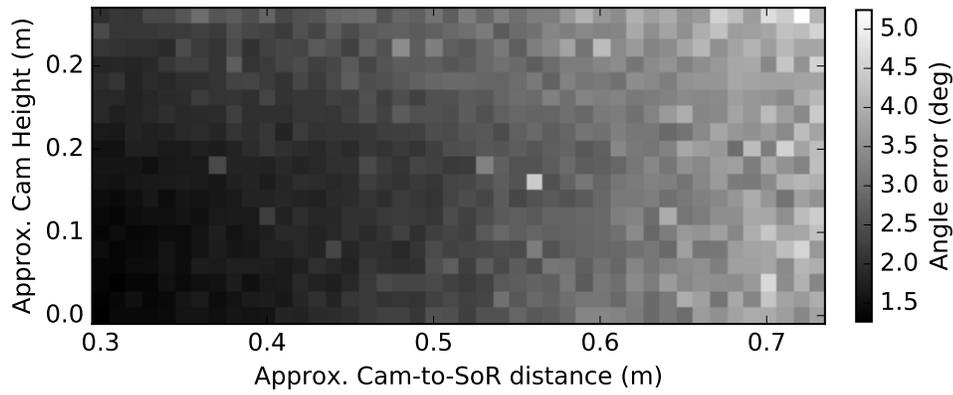
The imaged view-plane dataset is used in conjunction with the synthetic view-plane dataset in order to evaluate 67,081 view pairs for each of the 11 sensor noise models, for a total of 737,891 samples. The mean translation and axis-angle errors are presented for each sensor model in Figure 6.2. Performance gracefully degrades between sensor models Synth-0 and Synth-4, with a maximum translation error of 4.6mm and axis-angle error of 2.1° . The real sensor model has similar errors of 3.1mm and 2.2° for translation and axis-angle errors, respectively, suggesting that the real sensor model is approximated by the Synth-4 sensor model. Translation and angular error spike dramatically starting at the Synth-5 sensor model, exceeding 25mm translation error and 9° angle error. This jump mirrors a notable increase in symmetric error at the Synth-5 sensor model.

6.5.2. Effect of camera position and baseline

Triangulation requires two views, introducing the potential for the baseline between two views to affect accuracy independently from the individual camera positions. To explore this increased parameter space, 67,081 view pairs are evaluated for 49 SoR models, yielding



(a) Translation error by camera height and depth



(b) Axis-angle error by camera height and depth

Figure 6.6: Triangulated pose translation and axis estimation errors by camera depth and height. Pairs of views are selected with varying camera heights and depths. The recovered SoR pose is described by its central axis direction and the 2-DoF translation encoding its bearing and distance from the camera. These values are compared with the groundtruth pose locations and aggregated by closest camera height and depth (\hat{d}, \hat{h}) in the two corresponding mean error plots (a) and (b).

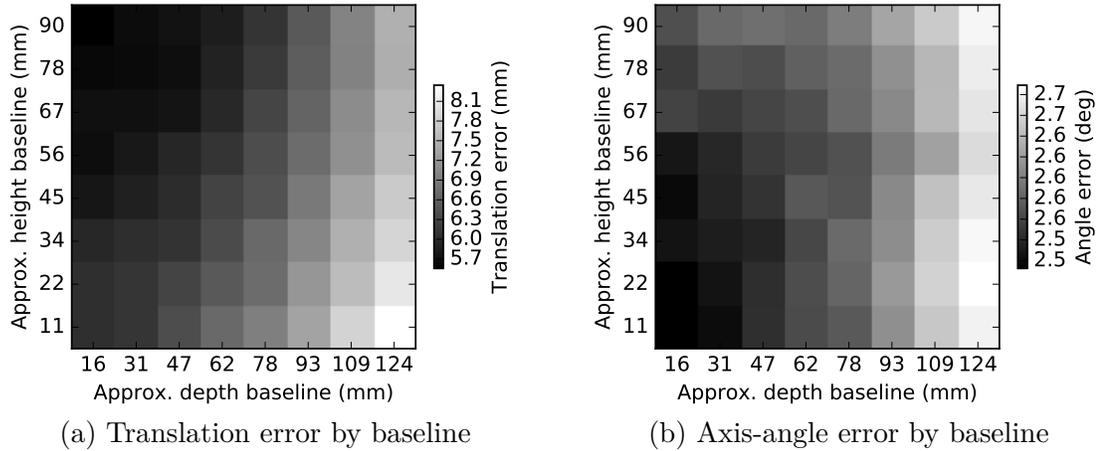


Figure 6.7: Triangulated pose translation and axis estimation errors by baseline. Pairs of views are selected with baselines that vary both in camera height $\Delta\hat{h}$ and camera depth $\Delta\hat{d}$. The recovered SoR pose is described by its central axis direction and the 2-DoF translation encoding its bearing and distance from the camera. These values are compared with the groundtruth pose locations and aggregated in by the baseline pairs $(\Delta\hat{d}, \Delta\hat{h})$ in the two corresponding mean error plots (a) and (b).

a total of 3,286,969 samples.

Figure 6.6 presents the triangulation accuracy in terms of translation and axis-angle error, aggregated by the camera position of the closest view to the SoR. This captures the general influence of distance of the camera to the SoR origin. Following the trend that is seen with symmetric pose estimation, the triangulation accuracy tends to decrease as the camera recedes from the SoR. This is not surprising as the accuracy of triangulation depends almost entirely on the prerequisite symmetric poses.

The effect of increased SoR-camera distance on triangulation accuracy is also seen by examining the triangulation error based on baseline, as shown in Figure 6.7. While there is no mathematical reason for a change in baseline as it is defined to influence triangulation error, accuracy tends to decrease as the baseline depth increases, yet is largely unaffected by baseline height. Since baseline errors are aggregated over all initial camera positions, larger baselines samples will encompass views of greater depths on average, and will be biased towards the increased error corresponding to such depths.

6.6. Summary

This chapter presented methods for 2D central-axis projection recovery and 3D central-axis triangulation that do not require bi-tangent points, visible cross sections, or a pseudo-symmetry assumption. The 2D central-axis projection is recovered using a two-point minimal problem that exploits the fact that the projection of an SoR is symmetric on the image sphere. The 2D central-axis projections of two views with a known relative camera transform are used to triangulate the 3D central axis. These methods were evaluated over a variety of noise conditions, poses, and baselines, and were shown to provide stable and accurate results in the presence of moderate noise, with pose errors that generally increase with increasing camera-to-SoR distance and increasing baseline depth.

7

One-point minimal correspondence problem for absolute pose

The pose, generatrix, and occluding contours of a surface of revolution are so strongly geometrically related that any two of these three entities contain sufficient information to recover the remaining one. The generatrix is recovered in the previous chapter using the absolute 5-DoF pose and the occluding contours; this chapter presents how the absolute pose is recovered as a function of the generatrix and occluding contours from a single view. All prior work on this problem (Dhome et al., 1990; Glachet et al., 1992) requires the presence of at least one visible cross-section from which the object attitude ϕ is recovered. Not only are cross-sections not required for pose recovery, as introduced in Phillips et al. (2016), but three of the camera pose parameters (attitude ϕ , depth \hat{d} , and height \hat{h}) can be recovered as a one-point minimal correspondence problem. The remaining two parameters of camera, roll and yaw, are encoded in the symmetric rotation ${}^s\mathbf{R}_r$ as recovered in Section 6.1.

7.1. One-point correspondence minimal problem

Referring back to the forward-projection Equation 4.23, an occluding contour point ${}^s\mathbf{x}$ and corresponding tangent line ${}^s\mathbf{l}$ in the representative image view are related to the canonical view by a rotation to the symmetric view of ${}^s\mathbf{R}_r$, followed by a rotation to the canonical view of ${}^c\mathbf{R}_s$. With a known estimate of ${}^s\mathbf{R}_r$, the transformed occluding contour point ${}^c\mathbf{x}$ and tangent ${}^c\mathbf{l}$ are dependent on the rotation ${}^c\mathbf{R}_s$, which is parameterized by attitude angle ϕ . Including the static rotation ${}^o\mathbf{R}_c$ from the canonical to object coordinates, a vector $(a_o, b_o, c_o)^\top$ in object coordinates is explicitly parameterized by ϕ and the vector $(a_s, b_s, c_s)^\top$

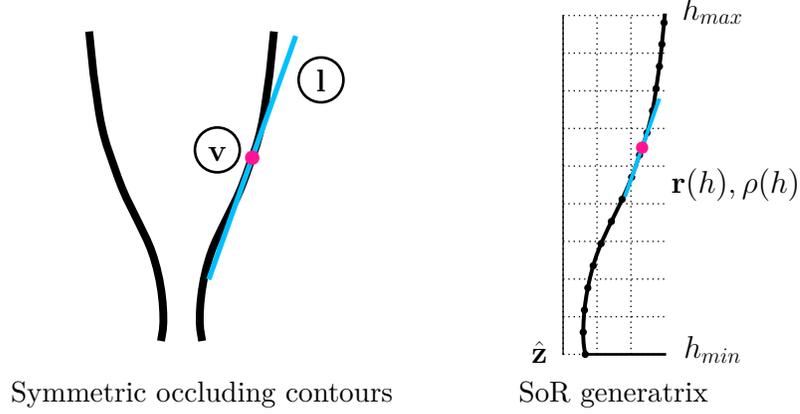


Figure 7.1: One-point correspondence for absolute pose. The correspondence between the symmetric occluding contour point \mathbf{v} with tangent line \mathbf{l} and generatrix point $(h, r(h), \rho(h))$ is sufficient to compute the three pose parameters (\hat{d}, \hat{h}, ϕ) .

in symmetric coordinates as

$$\begin{pmatrix} a_o \\ b_o \\ c_o \end{pmatrix} = \begin{pmatrix} -b_s \sin \phi - c_s \cos \phi \\ a_s \\ -b_s \cos \phi + c_s \sin \phi \end{pmatrix}. \quad (7.1)$$

Given one point correspondence of the form

$$(r, h, \rho) \leftrightarrow ({}^s\mathbf{x}, {}^s\mathbf{l}), \quad (7.2)$$

where (r, h, ρ) in the generatrix point corresponding to occluding contour point ${}^s\mathbf{x}$ and tangent line ${}^s\mathbf{l}$ in the symmetric view, then the camera pose parameter tuple (attitude ϕ , depth \hat{d} , height \hat{h}) can be recovered as a minimal problem. An example correspondence is illustrated in Figure 7.1.

7.1.1. Attitude recovery from generatrix derivative

Given an occluding contour point ${}^s\mathbf{x}$ with its tangent line ${}^s\mathbf{l}$ in the symmetric view, the attitude ϕ that relates the symmetric and canonical view can be expressed as a function of the point's corresponding generatrix derivative ρ . This is accomplished by the composition of two equations:

1. Equation 7.1 relates ${}^s\mathbf{l}$ to object coordinates ${}^o\mathbf{l}$ as a function of ϕ .
2. Equation 4.22 relates ${}^o\mathbf{l}$ to the corresponding generatrix derivative ρ .

Applying Equation 7.1 to ${}^s\mathbf{l}$ and massaging it with the Pythagorean identity,

$$1 = \cos^2 \phi + \sin^2 \phi,$$

allows the ρ -recovery parameterization from Equation 4.22 to be partially rewritten as

$$\rho = \frac{w_o}{\sqrt{u_s^2 + v_s^2 + w_s^2 - w_o^2}}, \quad (7.3)$$

which simplifies to

$$\rho = \frac{w_o}{\sqrt{-w_o^2 + 1}}, \quad (7.4)$$

by normalizing ${}^s\mathbf{l}$. Solving for w_o and representing it in terms of the rotated line ${}^s\mathbf{l}$ provides the relationship between ρ and ϕ ,

$$(\rho^{-2} + 1)^{-1/2} = -v_s \cos \phi + w_s \sin \phi, \quad (7.5)$$

expressed as a single sinusoid by the phase shift identity

$$(\rho^{-2} + 1)^{-1/2} = k \sin(\phi + \psi) \quad (7.6)$$

with constants

$$k = \sqrt{v_s^2 + w_s^2}, \quad \psi = \arctan2(w_s, v_s).$$

The solution is constrained to lie in front of the camera with

$$-\pi/2 \leq \phi \leq \pi/2.$$

The intuition behind this geometric constraint is that the generatrix tangent line corre-

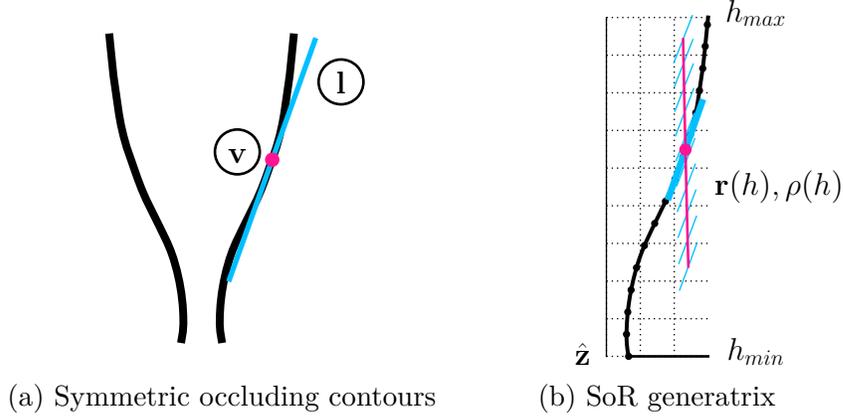


Figure 7.2: Remapped tangent lines as a function of attitude ϕ . For a fixed occluding contour point \mathbf{v} with tangent line \mathbf{l} (a), varying the attitude parameter ϕ results in a generator point $(h, r(h), \rho(h))$ that spans the generatrix space as shown in (b).

sponding to generatrix derivative ρ directly maps to the tangent line ${}^o\mathbf{l}$ in the canonical view, which is then remapped by a rotation of ϕ about the camera's x -axis to ${}^s\mathbf{l}$ in the symmetric view. The set of generatrix tangent lines that map to the same tangent line \mathbf{l}_s under varying values of ϕ is shown in Figure 7.2. The value of ϕ must verify that the resulting generatrix tangent line matches the known generatrix derivative ρ .

7.1.2. Translation recovery from generatrix point

Given an apparent contour point ${}^c\mathbf{x}$ and tangent line ${}^c\mathbf{l}$ in the canonical view, along with its corresponding generatrix point (r, h) , it is possible to solve for the representative camera depth \hat{d} and height \hat{h} .

The camera ray Equation 4.20 expresses a contour generator point in terms of ${}^c\mathbf{x}$ and ${}^c\mathbf{l}$ expressed in object space as:

$$\begin{pmatrix} \tilde{x}_o \\ \tilde{y}_o \\ \tilde{z}_o \end{pmatrix} = \mu \hat{d} \begin{pmatrix} x_o \\ y_o \\ z_o \end{pmatrix} + \begin{pmatrix} \hat{d} \\ 0 \\ \hat{h} \end{pmatrix}, \quad (7.7)$$

with

$${}^o\mathbf{x} = (x_o, y_o, z_o)^\top$$

$${}^o\mathbf{1} = (u_o, v_o, w_o)^\top$$

$$\mu = \frac{v}{uy - xv}.$$

Solving for \hat{d} using these equations and the known radius r , we have:

$$r^2 = \tilde{x}^2 + \tilde{y}^2 \tag{7.8}$$

$$r^2 = (\mu\hat{d}x + \hat{d})^2 + (\mu\hat{d}y)^2 \tag{7.9}$$

$$r^2 = (\mu x + 1)^2 \hat{d}^2 + (\mu + y)^2 \hat{d}^2 \tag{7.10}$$

$$r^2 = ((\mu x + 1)^2 + (\mu y)^2) \hat{d}^2 \tag{7.11}$$

$$\hat{d} = \sqrt{r^2 / ((\mu x + 1)^2 + (\mu y)^2)}, \tag{7.12}$$

and

$$\hat{h} = \tilde{z} - \mu\hat{d}z = h - \mu\hat{d}z, \tag{7.13}$$

for \hat{h} using \hat{d} and the known height h .

Intuitively, tangency information provides the attitude, but not does provide information about the camera's height or depth with respect to the object. The correspondence between generatrix point (r, h) and image point ${}^c\mathbf{x}$ constrains the scale of the contour generator by r , as well as its height in the object's coordinate system by h .

7.2. Absolute pose recovery evaluation

The absolute pose of an imaged SoR with a known generatrix is recovered using a one-point minimal correspondence problem. Given a hypothesized correspondence between an apparent contour point and a generatrix point, a corresponding absolute pose hypothesis is directly computed. Such a pose hypothesis encodes three degrees of freedom, the represen-

tative camera depth \hat{d} and height \hat{h} as well as the SoR central axis attitude ϕ . Two natural metrics of estimated pose accuracy are therefore the translation and attitude error.

The performance of the absolute pose estimate technique is evaluated with respect to three parameters: sensor noise, camera position, and SoR shape. This is achieved by the use of both real and synthetic datasets defined over a plane of camera view locations.

7.2.1. Ranked hypothesis generation

The hypothesis generation and scoring procedure discussed in Section 6.1.2 is used to create a ranked list of absolute pose hypotheses. This procedure requires a space of potentially corresponding input points for use in the minimal problem. In Phillips et al. (2016), this space is initialized by using a Procrustes-like (Hurley and Cattell, 1962) algorithm to yield an approximate apparent contour alignment. To isolate any potential failure in this sub-algorithm, the correspondence search space is initialized as a dilation from the groundtruth correspondence set to produce a space of similar size that is guaranteed to contain approximate correspondences.

7.2.2. Rank and recall analysis

Since absolute pose recovery is a hypothesis generation procedure (Section 6.1.2) that produces a ranked list of hypotheses, there is an inherent hypothesis selection problem that needs to be addressed to evaluate the technique. The reconstruction error corresponding to a hypothesized pose is used to rank all pose hypotheses with a sufficient number of inlying correspondences. This provides a reasonable baseline selection mechanism that allows the minimal problem to be examined. The estimation error of the best model that is present below a specified rank is a useful metric for evaluating the overall quality and recall of the hypotheses produced by the estimation procedure. The closer the error fall-off is towards rank 1, the better the recall, while the lower the trailing error values, the higher quality the estimation. Figure 7.3 illustrates such a rank analysis over the 900 views of the synthetic view-plane database parameterized with noise model Synth-2 and generatrix SoR-34. The

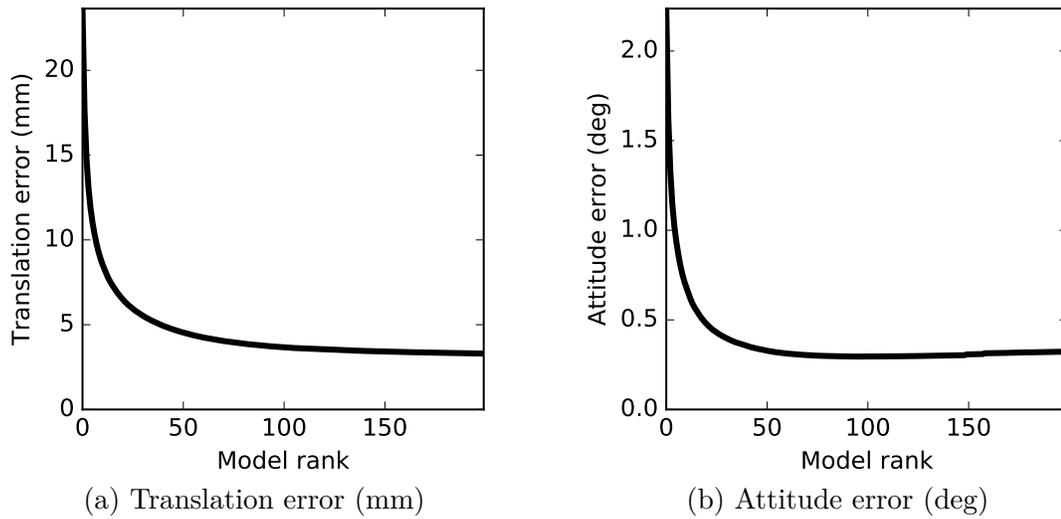


Figure 7.3: Absolute pose translation and attitude errors by model rank. Absolute pose hypotheses are ranked by geometric reconstruction error w.r.t. the target SoR generatrix model. The estimation error of the best model that is present below a specified rank is a useful metric for evaluating the overall quality and recall of the hypotheses produced by an estimation procedure. The closer the error fall-off is towards rank 1, the better the recall. The lower the trailing error values, the higher quality the estimation. The best model is seen to be contained within the first 50 hypotheses, with an average translation error (a) of 4.5mm and attitude error (b) of 0.32 degrees for the Synth-2 noise model and SoR-34.

best model is seen to be contained within the first 50 ranked hypotheses, with an average translation error (Figure 7.3a) of 4.2mm and attitude error (Figure 7.3b) of 0.32 degrees.

7.2.3. Effect of sensor noise model

The imaged view-plane dataset is used in conjunction with the synthetic view-plane dataset to evaluate 900 views of the generatrix SoR-34 for each of the 11 sensor noise models, yielding 9,900 samples in total. Table 7.1 presents the absolute pose errors by model rank for each of the sensor models. Pose errors are expressed in terms of translation and attitude error, with translation error further reported in terms of its depth and height components. At rank 20, the imaged dataset has approximately the same performance as the Synth-2 noise model, with an average translation of 4.1mm as compared to 4.3mm. By rank 50, all sensor models have mostly converged with translation errors growing gradually from 3.3mm for Synth-1 to 5.2mm for Synth-9. The real sensor dataset has 1.9mm mean translation error as opposed to 0.8mm for the least noisy model, Synth-0.

7.2.4. Effect of camera position

The synthetic view-plane dataset consists of 900 views and is parameterized with the Synth-2 noise model for all 49 SoR models to yield 44,100 sample points. The estimation errors of these views are evaluated with respect to the camera positions parameterized by depth \hat{d} and height \hat{h} . The effect of camera position on pose estimation accuracy is shown in Figure 7.4. Errors are presented in terms of translation and attitude error, with translation further reported in terms of its depth and height components. Over all metrics, absolute pose accuracy is the highest near the SoR and decreases with increased distance. A likely explanation for this behavior is that the noise-to-signal ratio increases as distance increases and apparent object size decreases. The absolute pose minimal-problem requires accuracy in both apparent contour point localization and tangent estimation, a condition that becomes harder to meet with diminished effective resolution due to distance.

Noise model		Error by model rank							
		1 st	5 th	10 th	20 th	30 th	40 th	50 th	100 th
Translation (mm)	Real	19.8	10.6	7.2	4.1	2.9	2.3	1.9	1.3
	Synth-0	3.0	1.5	1.1	0.9	0.9	0.8	0.8	0.8
	Synth-1	7.5	4.7	4.2	3.8	3.6	3.4	3.3	2.3
	Synth-2	8.8	5.4	4.7	4.3	4.0	3.8	3.6	2.4
	Synth-3	9.8	6.0	5.4	5.0	4.7	4.4	4.1	3.0
	Synth-4	13.0	7.1	6.0	5.5	5.1	4.8	4.4	3.5
	Synth-5	16.8	7.7	6.7	5.9	5.4	4.9	4.6	3.6
	Synth-6	14.4	8.5	7.3	6.3	5.6	5.1	4.7	4.0
	Synth-7	16.8	9.3	7.8	6.5	5.6	4.9	4.6	3.9
	Synth-8	18.4	10.8	8.6	6.8	5.9	5.3	4.8	3.9
Synth-9	18.9	10.2	8.1	6.8	6.2	5.6	5.2	4.2	
Attitude (deg)	Real	2.01	1.07	0.73	0.41	0.28	0.21	0.17	0.12
	Synth-0	0.28	0.13	0.09	0.08	0.07	0.07	0.07	0.07
	Synth-1	0.54	0.22	0.15	0.13	0.13	0.13	0.13	0.13
	Synth-2	0.68	0.26	0.19	0.16	0.16	0.16	0.16	0.18
	Synth-3	0.76	0.29	0.23	0.21	0.20	0.19	0.20	0.20
	Synth-4	1.08	0.39	0.28	0.24	0.22	0.22	0.23	0.23
	Synth-5	1.40	0.43	0.32	0.27	0.26	0.24	0.24	0.24
	Synth-6	1.13	0.47	0.37	0.32	0.30	0.28	0.28	0.28
	Synth-7	1.34	1.16	0.84	0.96	0.53	0.50	0.30	0.30
	Synth-8	1.78	0.97	0.75	0.62	0.58	0.56	0.55	0.53
Synth-9	1.69	0.91	0.52	0.43	0.61	0.58	0.56	0.54	
Depth (mm)	Real	3.5	2.2	1.7	1.4	1.2	1.1	1.0	0.9
	Synth-0	0.4	0.3	0.4	0.4	0.4	0.4	0.4	0.5
	Synth-1	4.2	3.8	3.6	3.3	3.2	3.0	2.9	1.8
	Synth-2	4.4	4.2	4.0	3.7	3.5	3.3	3.0	1.7
	Synth-3	4.8	4.6	4.5	4.3	4.0	3.7	3.3	2.3
	Synth-4	5.1	5.0	4.9	4.6	4.3	4.0	3.6	2.7
	Synth-5	5.7	5.4	5.2	4.9	4.4	4.0	3.6	2.6
	Synth-6	6.2	5.9	5.6	4.9	4.3	3.9	3.5	3.0
	Synth-7	6.1	5.6	5.1	4.6	4.0	3.6	3.3	2.8
	Synth-8	5.7	5.4	5.1	4.5	4.0	3.7	3.3	2.7
Synth-9	5.9	5.5	5.2	4.7	4.3	4.0	3.8	3.0	
Height (mm)	Real	19.3	10.1	6.6	3.5	2.3	1.7	1.3	0.8
	Synth-0	2.9	1.4	1.0	0.8	0.7	0.7	0.6	0.6
	Synth-1	5.7	2.3	1.7	1.4	1.3	1.3	1.3	1.2
	Synth-2	7.1	2.8	2.0	1.7	1.6	1.5	1.5	1.4
	Synth-3	7.8	3.0	2.4	2.1	1.9	1.9	1.8	1.5
	Synth-4	11.1	4.1	2.9	2.3	2.2	2.1	2.0	1.7
	Synth-5	14.5	4.5	3.3	2.7	2.4	2.3	2.2	1.9
	Synth-6	11.9	5.0	3.9	3.1	2.9	2.6	2.5	2.1
	Synth-7	14.3	6.1	4.8	3.7	3.1	2.7	2.5	2.2
	Synth-8	16.4	8.1	5.7	4.1	3.5	3.1	2.8	2.3
Synth-9	16.9	7.4	5.2	4.0	3.6	3.3	2.9	2.3	

Table 7.1: Absolute pose errors by hypothesis rank and noise model. The estimation errors of the best model present below the specified ranks are presented for the real and synthetic dataset with 10 noise models for SoR-34. Errors are reported in terms of translation and attitude error, with translation further presented in terms of its depth and height components.

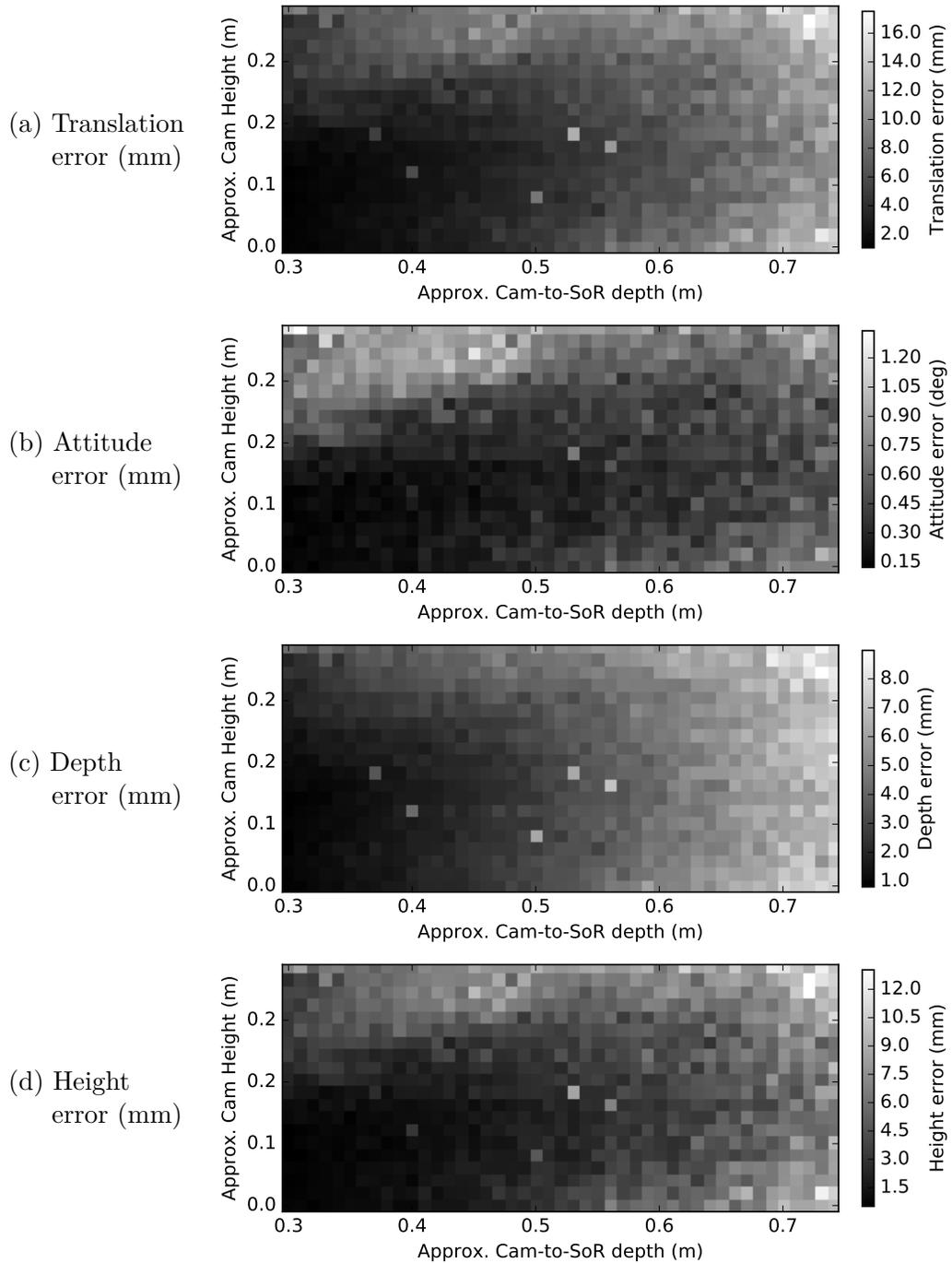


Figure 7.4: Absolute pose estimation errors by camera depth and height. The camera center is parameterized by a point (\hat{d}, \hat{h}) , representing the depth and height sampled from a plane of views. Each point represents the pose error computed over the synthetic views with noise model Synth-2 for all 49 SoR models. Pose error is reported in terms of (a) translation and (b) attitude. Translation error is further presented in terms of its (c) depth and (d) height components.

7.2.5. Effect of SoR shape

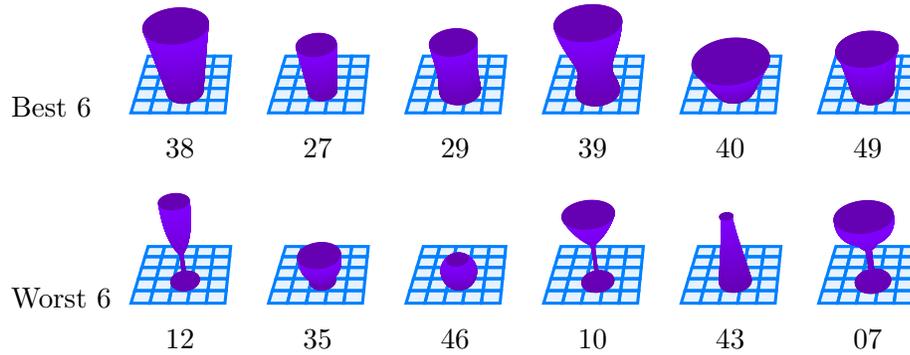


Figure 7.5: SoR models ranked by mean absolute pose estimate accuracy. The SoR models are ranked in order of increasing mean absolute pose translation error at rank 20. The six best and worst SoR models are presented in order of increasing error along with their associated identification numbers.

A qualitative analysis of the effect of SoR shape on absolute pose estimation is performed by first ranking the SoRs by their average estimation error. Figure 7.5 shows the best and worst six SoR models based on their average estimation error at rank 20. One observation is that the best six have larger average radii than the worst six. Following the trend that objects with smaller apparent sizes have greater pose estimation errors, this observation seems a plausible explanation for the improved performance on larger objects.

7.3. Summary

This chapter presented a method for absolute pose recovery that does not require bi-tangent points or visible cross sections. The absolute pose is recovered using a one-point minimal correspondence problem that directly relates the imaged points and tangents to the generatrix height, radius and derivative. It exploits this information and the forward-projection decomposition to recover the SoR attitude and position. This method was evaluated over a variety of noise conditions, poses, and SoR shapes, and was shown to provide stable and accurate results in the presence of moderate noise, with pose errors that generally increase

with increasing camera-to-SoR distance.

8

Two-point minimal correspondence problem for structure from motion

A classic result of Structure from Motion (SfM) allows the reconstruction of smooth algebraic surfaces from their 2D apparent contours using a Euclidean parameterization of the back-projected image rays (Giblin and Weiss, 1987; Cipolla, 1991; Cipolla and Blake, 1992). This chapter presents a two-point minimal correspondence problem for SfM for the case of an unknown model, allowing for the recovery of the relative poses of the SoR and two cameras, modulo scale and a single axis motion about the SoRs axis. With this pose information, the shape of the SoR is recovered up to an unknown scale.

Chapter 2 shows how pose and shape recovery is possible from the apparent contours seen from two views that have a known relative transform. The relative transform encodes both translation and attitude information as both aspects are uniquely determined by a triangulation of back-projected 2D SoR central-axes. Without the relative transform, these two parameters need to be recovered using the geometric constraints relating the SoR to its apparent contours.

Constraints of this nature are exploited in Chapter 7 to recover the translation and attitude of an SoR from the apparent contours of a single image given a known SoR generatrix. The two-view SfM formulation is an extension of the single-view case, where a second apparent contour view takes the place of a known generatrix. This substitution introduces an additional unknown attitude parameter that must be recovered, as well as a scale ambiguity that is inherent to SfM. The two unknown attitude parameters versus one explains why the SfM minimal problem requires two point correspondences instead of one, as is the case for

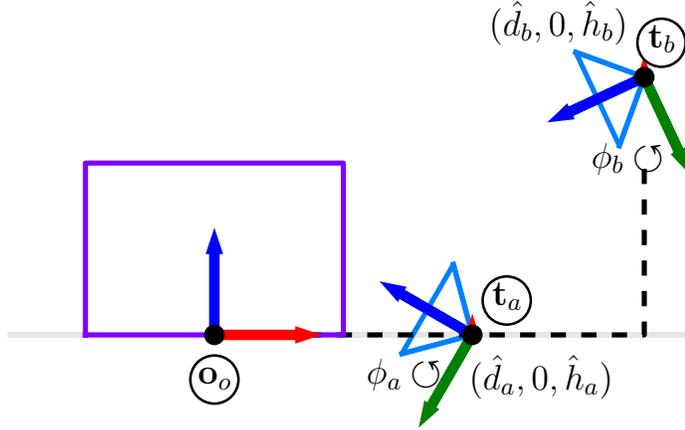


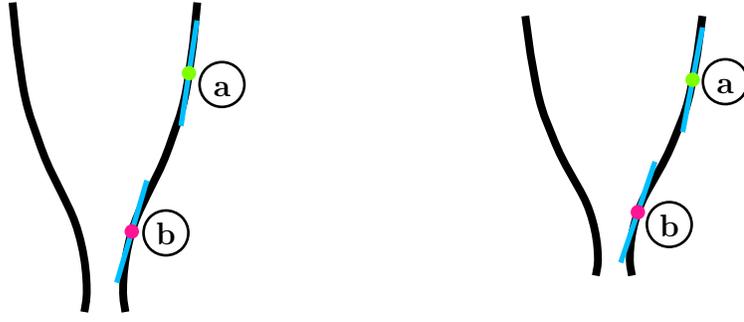
Figure 8.1: SfM two-view system pose parameterization. The position of each camera is specified in the representative pose as the distance \hat{d} along the xy -plane and the height \hat{h} above it. Due to the scale ambiguity of SfM and the arbitrary choice of height origin, the four parameters of (\hat{d}_a, \hat{h}_a) and (\hat{d}_b, \hat{h}_b) are encoded more compactly as $(\Delta d, \Delta h)$. The attitude of the two cameras are represented by the parameters (ϕ_a, ϕ_b) , for a total of four unknowns.

absolute pose recovery. In fact, once the attitude of either view is known, the SfM problem reduces to the absolute pose problem with an unknown scale.

8.1. Two-view system pose parameterization

The 5-DoF SoR pose is decomposed following the forward-projection Equation 4.23 as the symmetric rotation ${}^s\mathbf{R}_r$ and parameter tuple (ϕ, \hat{d}, \hat{h}) , leading to the symmetry rotation minimal problem of Chapter 6 and the absolute pose minimal problem of Chapter 7. The number of parameters required to describe the system increases with the addition of another view to form a view pair, however it does not quite double. The symmetric rotation encoding the camera pitch and yaw is estimated for each image separately, contributing 4-DoF, however the remaining elements of the system configuration can be described by a 4-tuple parameter, totaling 8-DoF.

Assuming two parameter tuples are required, one for each camera a and b , a 6-tuple $(\phi_a, \hat{d}_a, \hat{h}_a, \phi_b, \hat{d}_b, \hat{h}_b)$ would be required to describe the camera poses. However, due to the scale ambiguity inherent in SfM as well as the freedom to choose the height offset



View 1 symmetric occluding contours View 2 Symmetric occluding contours

Figure 8.2: Two-point minimal problem for structure from motion. Given occluding contour point correspondences a and b between the two views, the pose of both SoRs can be recovered up to an unknown scale and arbitrary height offset.

for the reconstructed generatrix, a reduction of 2-DoF allows for a 4-tuple of the form $(\phi_a, \phi_b, \Delta d, \Delta h)$. The two attitude parameters are irreducible, although may also be represented as ϕ_a and $\Delta\phi = \phi_b - \phi_a$ for mathematical convenience. Choosing the system scale to be such that $\hat{d}_a = 1$, then $\hat{d}_b = 1 + \Delta d$ and the first DoF reduction is seen. The second reduction occurs by choosing $\hat{h}_a = 0$, with $\hat{h}_b = 0 + \Delta h$.

8.2. Two-point minimal correspondence problem

8.2.1. Dual attitude recovery

Recall from the absolute pose formulation in Section 7.1.1 that the attitude of a single SoR view with known generatrix can be recovered using one correspondence of the form

$$(r, h, \rho) \leftrightarrow (\mathbf{x}_s, \mathbf{l}_s)$$

between the generatrix and the occluding contour in conjunction with Equation 7.6,

$$(\rho^{-2} + 1)^{-1/2} = k \sin(\phi + \psi).$$

Here the left hand side encodes the derivative from the (known) generatrix, and the right hand side encodes the point and tangent information from the occluding contour correspondence. In an SfM formulation, the generatrix is unknown and therefore the required derivative ρ is not directly measurable.

In this two-view SfM formulation, the views a and b contain occluding contours of an SoR of the same generatrix. If one occluding contour point correspondence,

$$(\mathbf{x}_{a1}, \mathbf{l}_{a1}) \leftrightarrow (\mathbf{x}_{b1}, \mathbf{l}_{b1}),$$

is known, where correspondence implies the two points are generated by the same (unknown) generatrix point (r_1, h_1, ρ_1) , then Equation 7.6 becomes useful again. Duplicating the right hand side of the equation for each point produces the three-way equality

$$(\rho_1^{-2} + 1)^{-1/2} = k_{a1} \sin(\phi_a + \psi_{a1}) = k_{b1} \sin(\phi_b + \psi_{b1}), \quad (8.1)$$

which constrains the values of ϕ_a and ϕ_b even though the value of ρ_1 is unknown. Equation 8.1 is one equation with two unknowns, therefore another equation is required to solve for both ϕ_a and ϕ_b . Assuming a second contour point correspondence

$$(\mathbf{x}_{a2}, \mathbf{l}_{a2}) \leftrightarrow (\mathbf{x}_{b2}, \mathbf{l}_{b2})$$

provides the system of equations

$$\begin{aligned} k_{a1} \sin(\phi_a + \psi_{a1}) &= k_{b1} \sin(\phi_b + \psi_{b1}) \\ k_{a2} \sin(\phi_a + \psi_{a2}) &= k_{b2} \sin(\phi_b + \psi_{b2}), \end{aligned} \quad (8.2)$$

which fully constrains ϕ_a and ϕ_b .

Solving the system of equations 8.2 for ϕ_a yields the quadratic form

$$(b^2 - 1)x^2 + 2abx + (a^2 + c - 1) = 0,$$

with

$$x = \cot(\phi_a)$$

and constants (a, b, c) encoding the phase shifts and amplitudes of the system. The attitude ϕ_a is recovered as $\arctan(x^{-1})$ and ϕ_b is recovered by back-substitution; see the Appendix C.1 for the full derivation.

8.2.2. Relative translation recovery

Due to the scale ambiguity inherent to SfM and the arbitrary choice of generatrix height offset, the 4-DoF of the two representative camera centers $(\hat{d}_a, 0, \hat{h}_a)$ and $(\hat{d}_b, 0, \hat{h}_b)$ are reduced by 2-DoF to the relative camera center translation $(\Delta_d, 0, \Delta_h)$. Under this parameterization, the representative camera center of camera a is fixed as

$$\mathbf{t}_a^o = (\hat{d}_a, 0, \hat{h}_a)^\top = (1, 0, 0)^\top,$$

while the center of camera b is expressed by the relative translation

$$\mathbf{t}_b^o = \mathbf{t}_a^o + (\Delta_d, 0, \Delta_h)^\top.$$

A contour generator point is parameterized by the linear form

$$\begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{pmatrix} = \begin{pmatrix} (\mu x + 1)\hat{d} \\ \mu y \hat{d} \\ \mu z \hat{d} + \hat{h} \end{pmatrix} \quad (4.20)$$

as a function of image data (x, y, z, μ) and representative camera center $(\hat{d}, 0, \hat{h})^\top$. Substituting in the above forms for camera centers \mathbf{t}_a and \mathbf{t}_b into Equation 4.20 yields two

parameterized contour generator points,

$$\begin{pmatrix} \tilde{x}_a \\ \tilde{y}_a \\ \tilde{z}_a \end{pmatrix} = \begin{pmatrix} (\mu_a x_a + 1) \\ \mu_a y_a \\ \mu_a z_a \end{pmatrix}, \quad \begin{pmatrix} \tilde{x}_b \\ \tilde{y}_b \\ \tilde{z}_b \end{pmatrix} = \begin{pmatrix} (\mu_b x_b + 1)(1 + \Delta_d) \\ \mu_b y_b(1 + \Delta_d) \\ \mu_b z_b(1 + \Delta_d) + \Delta_h \end{pmatrix}, \quad (8.3)$$

with the first contour point \mathbf{x}_a fixed and the second \mathbf{x}_b dependent on the camera center translation $(\Delta_d, 0, \Delta_h)^\top$. Assuming points \mathbf{x}_a and \mathbf{x}_b correspond to the same generatrix point (r, \tilde{z}) , these two points provide a system of equations that is sufficient to recover the translation parameters as

$$\begin{aligned} \Delta_d &= \sqrt{\frac{\tilde{x}_a^2 + \tilde{y}_a^2}{(\mu_b x_b + 1)^2 + (\mu_b y_b)^2}} - 1 \\ \Delta_h &= \tilde{z}_a - \mu_b z_b(1 + \Delta_d). \end{aligned} \quad (8.4)$$

See Appendix C.1.1 for the full derivation.

8.3. Structure-from-motion pose recovery evaluation

Structure from motion from two views of an SoR is achieved using a two-point minimal correspondence problem. Given two hypothesized correspondences between the two SoR views, a corresponding pose hypothesis and associated generatrix reconstruction is directly computed. Due to scale ambiguity and the arbitrary selection of a height reference point, such a pose hypothesis encodes four degrees of freedom. The first two are the representative camera depth \hat{d} and height \hat{h} of second view with the first view fixed arbitrarily at a depth of 1 and height of 0. The last two are the two attitudes (ϕ_a, ϕ_b) of the two views.

The scale and height offset of the two views is recovered using the groundtruth pose, allowing the second estimated view pose and its groundtruth pose to be directly compared. Two reasonable metrics of SfM pose accuracy are the translation error of the second view and the combined attitude errors of both views. The translation of the first view encodes two free degrees of freedom and is used to recover the scale and height offset of the second view

with respect to the groundtruth.

In the absolute pose minimal problem, the mean reconstruction error of the view with respect to a known generatrix is used to rank candidate pose hypotheses (Section 7.2.1). In the two-view SfM minimal problem, the mean reconstruction between the two views is used for this purpose.

8.3.1. Effect of sensor noise model

The imaged view-plane dataset is used in conjunction with the synthetic view-plane dataset to evaluate 17,152 views pairs of the generatrix SoR-34 for each of the 11 sensor noise models, yielding 188,672 samples in total. The effect of camera position on pose estimation accuracy is shown in Figure 8.1. Errors are presented in terms of translation and attitude error, with translation further reported in terms of its depth and height components.

By rank 50, all sensor models have mostly converged with translation errors growing gradually from 10.4mm for Synth-1 to 16.3mm for Synth-9. The real sensor dataset has 21.0mm mean translation error as opposed to 6.2mm for the least noise model, Synth-0. Height error contributes the most to the translation error, with all models having less than 3.5mm of depth error at rank 50, as opposed to a height errors in the range of 6.1–19.6mm. This is an expected phenomenon for objects viewed at a large distance along the depth axis, as a small error in estimated attitude translates to a comparatively large error in height relative to depth.

8.3.2. Effect of camera position and baseline

As SfM requires two views, the effect of the baseline between two views on estimation accuracy is examined independently from the individual camera positions. To explore this increased parameter space, 17,152 view pairs are evaluated for 49 SoR models with the Synth-2 noise model, yielding a total of 840,448 samples. Figure 8.3 presents the SfM pose accuracy aggregated by the camera position of the closest view to the SoR. Errors are

Noise model		Error by model rank							
		1 st	5 th	10 th	20 th	30 th	40 th	50 th	100 th
Translation (mm)	Real	52.0	30.8	26.5	23.5	22.0	21.0	20.2	17.7
	Synth-0	25.8	6.8	5.9	6.1	6.1	6.2	6.2	6.2
	Synth-1	42.6	19.9	16.3	13.6	12.1	11.1	10.4	9.1
	Synth-2	45.2	20.8	17.1	14.5	13.1	12.1	11.5	9.9
	Synth-3	45.8	20.9	17.4	14.9	13.4	12.4	11.8	10.0
	Synth-4	45.1	20.3	16.8	14.6	13.4	12.5	11.9	10.4
	Synth-5	43.5	20.1	17.0	14.9	13.9	13.1	12.6	11.1
	Synth-6	43.0	19.9	17.1	15.4	14.5	13.7	13.2	11.9
	Synth-7	42.1	20.3	17.8	16.3	15.4	14.7	14.1	12.7
	Synth-8	41.6	21.1	18.9	17.5	16.6	15.9	15.4	13.9
Synth-9	40.6	21.6	19.7	18.4	17.5	16.8	16.3	15.0	
Attitude (deg)	Real	27.4	11.0	7.3	5.0	4.0	3.5	3.1	2.4
	Synth-0	14.8	2.7	1.6	1.1	1.0	0.9	0.9	0.8
	Synth-1	23.1	6.0	3.5	2.3	1.8	1.6	1.5	1.2
	Synth-2	24.5	6.4	3.7	2.4	2.0	1.7	1.6	1.3
	Synth-3	24.6	6.4	3.8	2.5	2.0	1.8	1.6	1.3
	Synth-4	24.0	6.3	3.7	2.4	2.0	1.8	1.6	1.3
	Synth-5	23.1	6.3	3.8	2.5	2.1	1.8	1.7	1.4
	Synth-6	23.0	6.4	3.9	2.6	2.2	1.9	1.8	1.5
	Synth-7	22.5	6.4	4.0	2.7	2.3	2.1	1.9	1.6
	Synth-8	22.2	6.5	4.1	2.9	2.4	2.2	2.1	1.8
Synth-9	21.7	6.6	4.3	3.0	2.6	2.3	2.2	1.9	
Depth (mm)	Real	10.1	6.7	5.2	4.2	3.8	3.5	3.4	3.2
	Synth-0	1.0	0.8	0.7	0.8	0.8	0.9	0.9	1.0
	Synth-1	4.3	3.3	2.8	2.5	2.3	2.2	2.1	1.9
	Synth-2	4.2	3.4	3.1	2.8	2.6	2.4	2.3	2.1
	Synth-3	3.7	3.3	3.0	2.8	2.6	2.5	2.3	2.1
	Synth-4	3.2	3.0	2.9	2.7	2.6	2.4	2.3	2.2
	Synth-5	2.9	2.8	2.8	2.7	2.6	2.6	2.5	2.4
	Synth-6	2.9	2.9	2.8	2.8	2.7	2.7	2.6	2.5
	Synth-7	3.1	3.0	3.0	3.0	2.9	2.9	2.8	2.8
	Synth-8	3.4	3.2	3.2	3.2	3.1	3.1	3.1	3.0
Synth-9	3.5	3.3	3.3	3.3	3.3	3.3	3.2	3.2	
Height (mm)	Real	49.3	29.3	25.4	22.7	21.3	20.4	19.6	17.1
	Synth-0	25.6	6.6	5.8	5.9	6.0	6.1	6.1	6.0
	Synth-1	41.8	19.1	15.7	13.0	11.6	10.6	9.9	8.6
	Synth-2	44.5	20.0	16.4	13.9	12.5	11.5	10.9	9.3
	Synth-3	45.2	20.2	16.7	14.3	12.8	11.9	11.2	9.4
	Synth-4	44.7	19.7	16.2	14.0	12.9	12.0	11.3	9.8
	Synth-5	43.1	19.5	16.5	14.3	13.3	12.6	12.0	10.5
	Synth-6	42.6	19.3	16.5	14.8	13.9	13.1	12.6	11.2
	Synth-7	41.6	19.7	17.2	15.6	14.7	14.0	13.4	12.0
	Synth-8	41.1	20.5	18.2	16.9	15.9	15.2	14.6	13.2
Synth-9	40.0	20.9	19.0	17.7	16.8	16.1	15.5	14.2	

Table 8.1: Two-view structure-from-motion pose errors by hypothesis rank and noise model. The estimation errors of the best model present below the specified ranks are presented for the real and synthetic dataset with 10 noise models for SoR-34. Errors are reported in terms of translation and attitude error, with translation further presented in terms of its depth and height components.

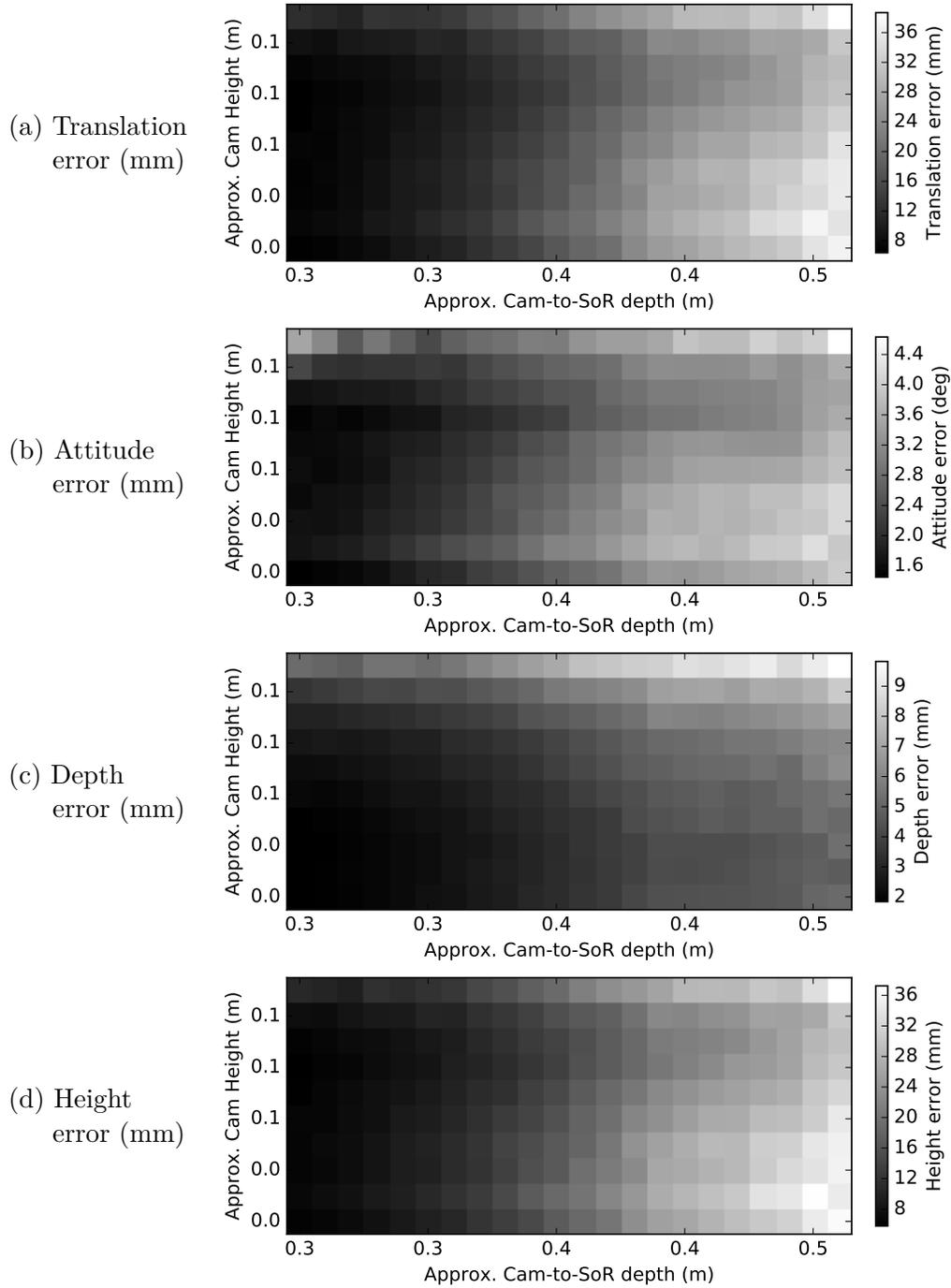


Figure 8.3: Two-view structure-frame-motion pose estimation errors by camera depth and height. Pairs of views are selected with varying camera depths and heights from a plane of views, with the first view at point (\hat{d}, \hat{h}) and second at point $(\hat{d} + \Delta\hat{d}, \hat{h} + \Delta\hat{h})$, with baseline $(\Delta\hat{d}, \Delta\hat{h})$. Each point represents the pose error computed over the synthetic views with noise model Synth-2 for all 49 SoR models. Pose error is reported in terms of (a) translation, (b) attitude, and translation (c) depth and (d) height components.

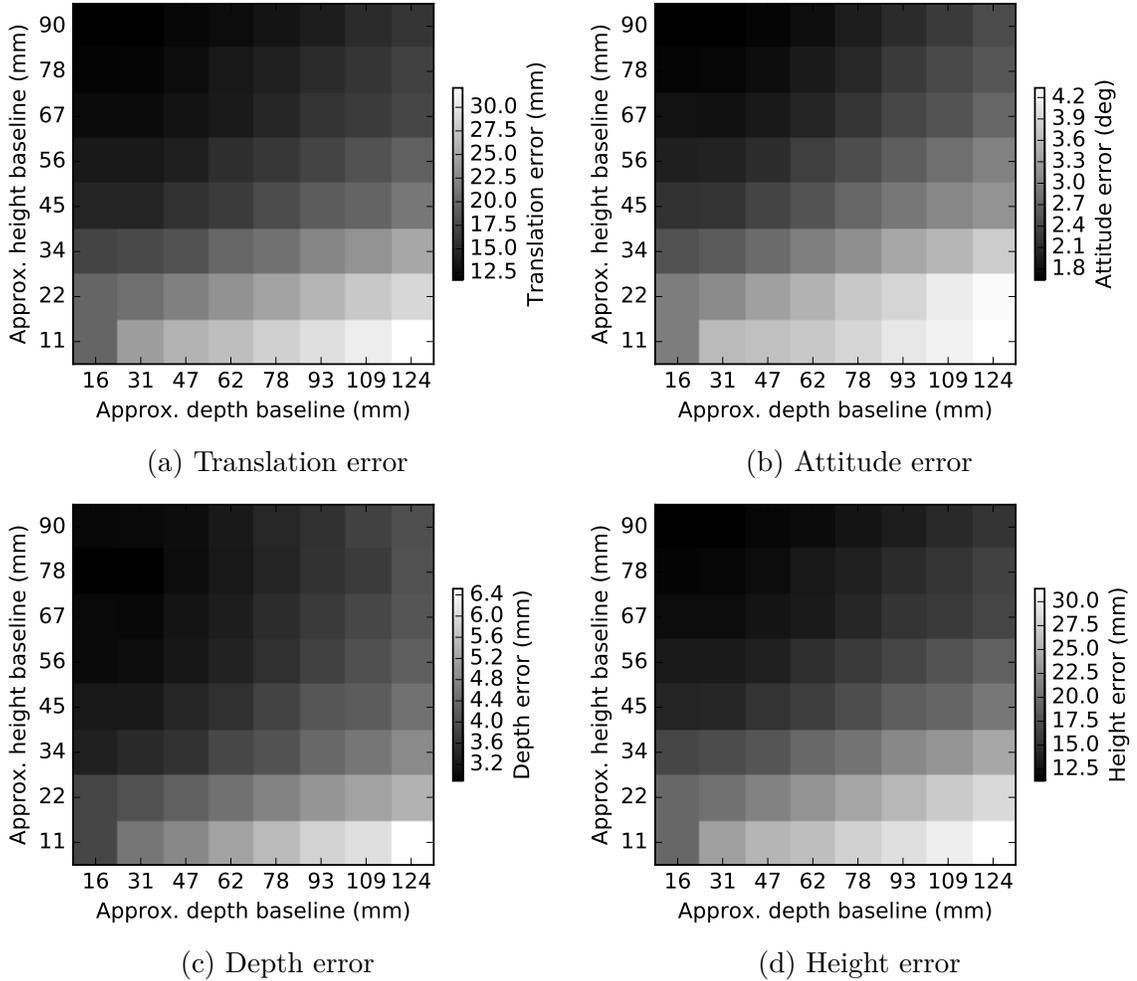


Figure 8.4: Two-view structure from motion pose estimation errors by depth and height baseline. Pairs of views are selected with varying camera depths and heights from a plane of views, with the first view at point (\hat{d}, \hat{h}) and second at point $(\hat{d} + \Delta\hat{d}, \hat{h} + \Delta\hat{h})$, where $(\Delta\hat{d}, \Delta\hat{h})$ is the inter-view baseline. Each point represents the error computed over synthetic views with noise model Synth-2 of all 49 SoR models. Errors are reported for a model rank of 100 in terms of (a) translation and (b) attitude. Translation error is further presented in terms of its (c) depth and (d) height components.

presented in terms of translation and attitude error, with translation further reported in terms of its depth and height components. The error graphs capture the general influence of camera distance to the SoR origin. SfM pose accuracy tends to decrease as the camera recedes from the SoR, following the intuition that as the apparent size of the SoR decreases, so does the fidelity of the apparent contour point and tangent estimates. Since SfM requires accurate contour points and tangents, any increase in signal noise will decrease performance.

Figure 8.4 shows the effect of the two-view baseline on estimated pose accuracy, reported in terms of translation, attitude, depth and height errors. SoR views with the largest baseline in height and smallest baseline in depth have the smallest estimation error across all metrics. This represents the combination of two competing factors, the first due to mathematical degeneracy, and the second due to signal degradation from increased distance as previously described. Structure from motion reaches a degenerate state if there is no motion; as the baseline distance approaches zero, the SfM solution becomes unstable.

For surfaces of revolution, the stability of a system is more accurately described by the motion of the contour generators (modulo single-axis motion) along the SoR's surface than the actual camera motion itself. As the camera recedes in depth, the contour generators converge to a planar shape, with little relative change as the camera continues to move. As the camera moves vertically however, there is typically a large variation in the contour generator position. This can be understood rather intuitively by considering the phenomenon of foreshortening. As one moves backwards from an object, the view angle remains relative constant, and the object mostly appears as the same shape, just smaller. As one moves vertically with respect to an object, their view angle must adjust to keep the object in view. Egocentrically, the object appears to tilt towards the viewer, with its shape distorted by foreshortening. Such an apparent shape change accompanies large changes in the contour generators, an advantage for system stability.

In this way, large baselines in height yield more stable results. Large baselines in depth are less effective in this regard, with the signal degradation due to lower effective resolution

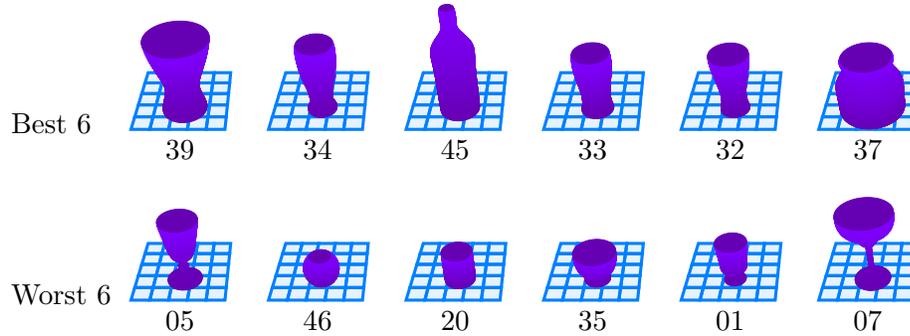


Figure 8.5: SoR models ranked by mean two-view absolute pose estimate accuracy. The SoR models are ranked in order of increasing mean absolute pose translation error. The six best and worst SoR models are presented in order of increasing error along with their associated identification numbers.

canceling out the relatively small motion in the contour generator.

8.3.3. Effect of SoR shape

A qualitative analysis of the effect of SoR shape on SfM pose estimation is performed by first ranking the SoRs by their average estimation error. Figure 8.5 shows the best and worst six SoR models based on their average estimation error at rank 20. One observation is that the best six have larger average radii than the worst six. Following the trend that objects with smaller apparent sizes have greater pose estimation errors, this observation seems a plausible explanation for the improved performance on larger objects.

8.4. Summary

This chapter presented a method for two-view SfM shape and pose recovery that does not require bi-tangent points, visible cross sections, or relative camera positions. The SfM pose is recovered using a two-point minimal correspondence problem that directly relates the imaged points and tangents from two views. It exploits this information and the forward-projection decomposition to recover the SoR attitudes and relative positions of these views. This method was evaluated over a variety of noise conditions, poses, SoR shapes, and baselines and was shown to provide stable and accurate results in the presence of moderate noise, with pose errors that generally increase with increasing camera-to-SoR distance and

increasing baseline depth.

9

Simultaneous n -view perceptual grouping and shape recovery

In previous chapters, the pose and shape of an imaged SoR is recovered assuming known apparent contours using the contour generator constraints for surfaces of revolution. In contrast, this chapter aims to recover the unknown apparent contour and generatrix assuming a known pose.

By assuming a known pose, the bijective mapping between the SoR generatrix and apparent contour can be treated as a geometric prior for simultaneous apparent contour perceptual grouping and generatrix recovery. This approach, as demonstrated in Phillips et al. (2016), has two key advantages over naïve perceptual grouping followed by shape reconstruction.

The first advantage is that the generatrices produced by the perceptual grouping technique are guaranteed to be geometrically plausible SoRs. Even perceptual grouping techniques that enforce smoothness may produce invalid generatrices of non-monotonically increasing height, with multiple radii values for the same height.

The second advantage comes from the fact that the grouping is performed in generatrix space. This property allows simultaneous bilateral contour grouping of a single image, aggregating edge evidence about both apparent contours into the same generatrix space, ensuring the resulting contours are symmetric on the view sphere. Furthermore, such evidence aggregation is not limited to a single view, as the information from an arbitrary n views can be leveraged to ameliorate the effects of noise, clutter, occlusions and low signal energy. In addition to an improvement in perceptual grouping performance, the runtime efficiency is significantly increased as the generatrix recovery procedure only needs to be



Figure 9.1: Three ideal SoR profiles.

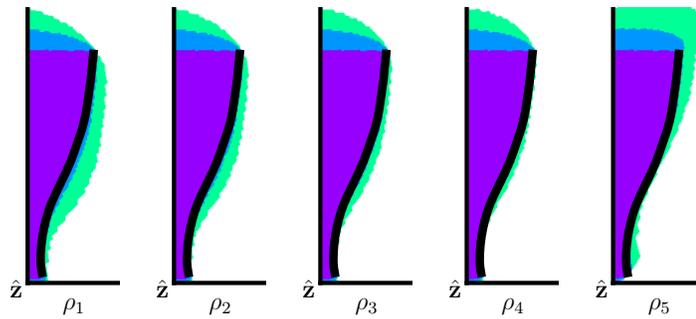


Figure 9.2: Three SoR image profiles mapped to generatrix space. This figure demonstrates how any number of SoRs can be mapped into generatrix space. The generatrix space is a volume, and presented here are generatrix space slices of constant derivative ρ . Multiple SoRs will only overlap perfectly at the (r, h, ρ) locations that match the generatrix values with which they were generated.

performed once over the aggregated energy map to simultaneously segment all n views. Three synthetic SoR profiles (Figure 9.1) are shown in Figure 9.2, illustrating how evidence from multiple images is mapped into the same generatrix space.

9.1. Apparent contour oriented edge response map

The bijection of Section 4.4.6 conceptually represents a mapping of the form

$$(u, v, \theta)^T \leftrightarrow (r, h, \rho)^T$$

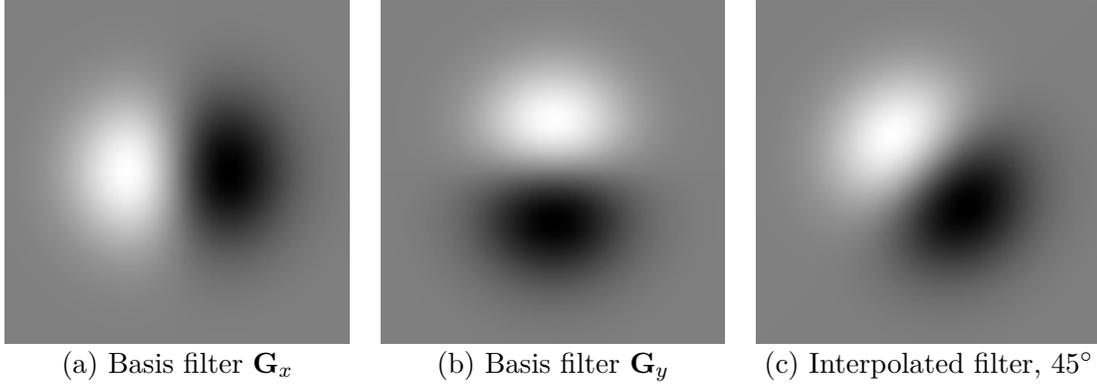


Figure 9.3: Steerable first-derivative Gaussian filter kernels. Interpolated filters (c) are created by the linear combination of the two base filters (a) and (b).

that relates the oriented image edge point (u, v) with angle θ to the generatrix point (r, h) with tangent derivative ρ . An oriented edge response function $\psi(u, v, \theta)$ is therefore required to apply the bijection and transfer the edge response from the image space to the generatrix space.

Steerable derivative filters (Freeman and Adelson, 1991) can be employed either to compute oriented edge responses directly from the source image, or to reaggregate the output of more sophisticated boundary detectors as a function of edge orientation. A steerable first-derivative Gaussian filter is presented in this chapter, which is best suited for step edges, however any oriented filter of the form $\psi(u, v, \theta)$ may be substituted as appropriate. Additionally, a higher-order odd derivative could be substituted for a “peakier” step edge response.

The steerable first-derivative Gaussian filter is defined by two kernels

$$\mathbf{G}_x(u, v) = \frac{-x}{\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}}, \quad \mathbf{G}_y(u, v) = \frac{-y}{\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}}, \quad (9.1)$$

representing base filters of orientations 0° and 90° that can be linearly combined to form a filter of any angle θ as

$$\mathbf{G}_\theta = \mathbf{G}_x \cos \theta + \mathbf{G}_y \sin \theta. \quad (9.2)$$

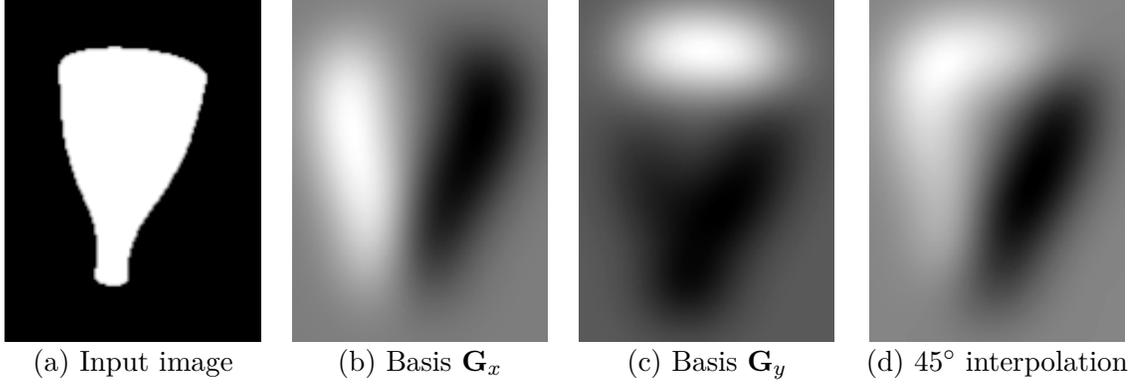


Figure 9.4: Steerable first-derivative Gaussian filter basis images and linear combination. The input image (a) is convolved with basis kernels \mathbf{G}_x and \mathbf{G}_y to produce basis images (b) and (c). Any direction can be interpolated (d) by the linear combination of these basis images.

Figure 9.3 illustrates these two basis kernels and an oriented kernel formed from their linear combination.

For image \mathbf{I} , the oriented edge response ψ for an angle θ is then defined as

$$\psi(\theta) = \mathbf{I} * \mathbf{G}_\theta = \mathbf{I} * (\mathbf{G}_x \cos \theta + \mathbf{G}_y \sin \theta). \quad (9.3)$$

Since convolution is a linear operator, the convolution of \mathbf{I} can be distributed as a weighted linear combination

$$\psi(\theta) = (\mathbf{I} * \mathbf{G}_x) \cos \theta + (\mathbf{I} * \mathbf{G}_y) \sin \theta \quad (9.4)$$

with $(\mathbf{I} * \mathbf{G}_x)$ and $(\mathbf{I} * \mathbf{G}_y)$ representing basis edge response images that only need to be calculated once per image. Such basis images are illustrated in Figure 9.4 along with an edge response resulting from their linear combination. In this way, the oriented edge response $\psi(u, v, \phi)$ can be computed for any image point, with any angle.

9.2. Generatrix contour space

Given the SoR pose, an oriented edge response map $\psi(u, v, \theta)$ in image space is transferred to the generatrix space $\gamma(r, h, \rho)$ by the back-projection half of the bijection between the ap-

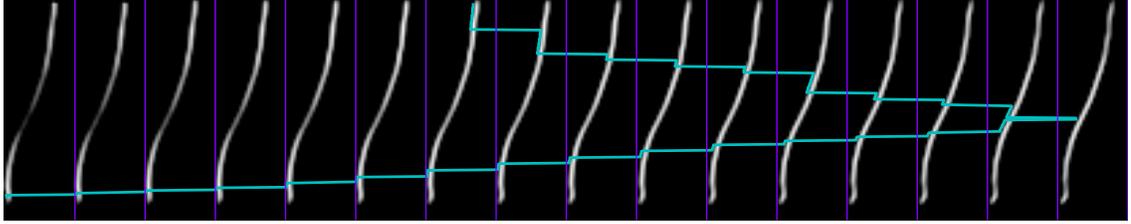


Figure 9.5: Optimal dynamic programming path through the generatrix parameter space. The generatrix space volume slices are stacked horizontally in increasing derivative value ρ from left to right, separated by vertical lines. The response of the mapped generatrix energy will be highest at points that match the generatrix value (r, h, ρ) to which it corresponds. The blue curve represents the generatrix that was used to create the imaged profile. Notice how the curve follows area of high energy.

parent contour and the generatrix (Section 4.4.6). Since the representation of the generatrix space is viewpoint independent, the energy of any number of arbitrary SoR views can be efficiently aggregated in this space. Leveraging edge and shape information from multiple views helps ameliorate the effects of noise, clutter, occlusions and response drop-outs.

The generatrix is recovered by optimizing the generatrix curve with respect to the contour energy back-projected from image space $\psi(u, v, \theta)$ into generatrix space $\gamma(r, h, \rho)$. The apparent contour grouping for each input image is determined by the forward projection of the generatrix into the image space using the known corresponding SoR pose. In this manner, perceptual grouping and generatrix recovery are performed simultaneously and are guaranteed to be geometrically consistent.

In practice, both the image space map $\psi(u, v, \theta)$ and generatrix space map $\gamma(r, h, \rho)$ are discretized bounded three-dimensional volumes. The volumes are typically examined in slices of constant θ for image space and ρ for generatrix space. Three such generatrix space slices are illustrated in Figure 9.2.

9.3. Dynamic program optimal subproblem for generatrix optimization

The generatrix edge response map $\gamma(r, h, \rho)$ is a discretized three-dimensional volume that provides the energy value for a generatrix curve point at height h with radius r and derivative

ρ . The “optimal” generatrix curve through this space is one that maximizes the total corresponding edge energy while satisfying the following properties:

1. \mathcal{C}^0 continuity: the curve must be single connected segment.
2. \mathcal{C}^1 continuity: the first derivative ρ must vary smoothly.
3. Low complexity: the curve must have no more than k inflexion points.

Figure 9.5 illustrates such a curve through generatrix space.

The selection of such an optimal generatrix from a discretized bounded generatrix space is an example of a combinatorial optimization problem. While the solution space is finite, a naïve brute force exploration of it tends to suffer from an exponential “combinatorial explosion”. A more intelligent approach recognizes and exploits the fact that the evaluation of two similar problems may depend on the solutions of a set of smaller overlapping subproblems, a property called *optimal substructure*.

Dynamic programming (Bellman, 1954) is such a combinatorial optimization technique that exploits a problem’s optimal substructure by recursively decomposing overlapping subproblems, solving them from bottom up. The smaller subproblem solutions are *memoized*, computed and stored, for repeated use in the evaluation of the larger subproblems. The effective use of memoization can have dramatic time complexity implications, often allowing an exponentially large combinatorial space to be searched in $\mathcal{O}(n^k)$ time, where k is the dimensionality of the space.

To apply dynamic programming to an optimization problem, two aspects must be defined. First is the definition of the optimal subproblem $\text{OPT}(\Omega)$ that exploits the optimal substructure. The function $\text{OPT}(\Omega)$ yields the optimal solution for parameters Ω by evaluating some number of subproblems $\text{OPT}(\Omega')$, where Ω' parameterizes a smaller instance of the subproblem. The second aspect, seemingly trivial yet important, is the traversal order and base cases for solving the larger optimal subproblems in terms of the smaller ones. These

elements inductively validate that it is feasible to solve the smaller subproblems necessary for solving the larger subproblems.

For ease of presentation, the three-dimensional optimal subproblem will be introduced first and then extended to four-dimensions to include the number of inflexion points.

9.3.1. *Generatrix shape representation*

To reduce the time and implementation complexity of the dynamic programming solution, the SoR generatrix is optimized as a piecewise-linear function represented by a path of points through the generatrix contour space. Any two adjacent points (h_a, r_a, ρ_a) and (h_b, r_b, ρ_b) with $h_a < h_b$ are constrained such that

$$h_b = h_a + \Delta_h \tag{9.5}$$

and

$$r_b = r_a + \rho_a \Delta_h, \tag{9.6}$$

ensuring that points are separated in height by the fixed algorithm parameter Δ_h , and that the line segment $(h_a, r_a) - (h_b, r_b)$ has slope ρ_a . The effect of these constraints is that the generatrix is guaranteed to have \mathcal{C}^0 continuity.

\mathcal{C}^1 continuity is approximated by bounding the discrete second derivative between adjacent points as

$$|\rho_a - \rho_b| < \Delta_\rho, \tag{9.7}$$

where Δ_ρ is a fixed algorithm parameter. The number k of curve inflection points is used to curve shape complexity, acting as a regularization parameter.

Once the optimal generatrix path is recovered, the piecewise linear approximation is smoothed by a corner-cutting subdivision curve scheme (Chaikin, 1974) to achieve \mathcal{C}^1 continuity.

Given a specific generatrix space discretization γ and algorithm parameters Δ_h and Δ_ρ , the

adjacency set function $\mathcal{A}(h, r, \rho)$ yields the set of all points $\{(h', r', \rho') \in \gamma \wedge h' < h\}$ that are adjacent to (h, r, ρ) while respecting the adjacency constraints of equations 9.5, 9.6 and 9.7.

9.3.2. Three-dimensional optimal subproblem

The three-dimensional optimal problem, $\text{OPT}(h, r, \rho)$, is stated to be the optimal generatrix composed of adjacent generatrix points in γ that extend from the lowest height h_{min} up to the maximum point with height h , radius r , and derivative ρ . Using this optimal problem, the optimal solution is defined recursively as

$$\text{OPT}(h, r, \rho) = \max_{(h', r', \rho') \in \mathcal{A}(h, r, \rho)} [\text{OPT}(h', r', \rho') + \psi(r, h, \rho)f(\rho, \rho')], \quad (9.8)$$

where the function $f(\rho, \rho')$ can be used to control the algorithm's preference for smooth generatrix second-derivatives. The optimal subproblem solution $\text{OPT}(h', r', \rho')$ on the right hand side of Equation 9.8 is extended to include point (h, r, ρ) by adding the energy along the line $(h', r', \rho') - (h, r, \rho)$ that is encoded in the function $\psi(r, h, \rho)$. All optimal subproblems that are adjacent to (h, r, ρ) are represented by the constrained adjacency set function $\mathcal{A}(h, r, \rho)$, and are therefore evaluated in determining the optimal solution for (h, r, ρ) .

Showing the feasibility of subproblem memoization, Equation 9.8 can be evaluated by memoizing in order of increasing height values with the base case defined as $\text{OPT}(0, \cdot, \cdot) = 0$. From Equation 9.5, h' is strictly less than h , meaning $\text{OPT}(h', r', \rho)$ will always be computed prior to $\text{OPT}(h, r, \rho)$, as is required for feasibility.

9.3.3. Four-dimensional optimal subproblem with inflection constraint

To constrain generatrix complexity by limiting the number of inflection points, another dimension is required in the optimal subproblem. The new optimum problem $\text{OPT}(k, h, r, \rho)$ is the same problem statement as $\text{OPT}(h, r, \rho)$, except that the generatrix must contain (at

most) k inflection points. The optimal problem recursive definition,

$$\text{OPT}(k, h, r, \rho) = \underset{(h', r', \rho') \in \mathcal{A}(k, h, r, \rho)}{\text{argmax}} \begin{cases} \text{OPT}(k, h - 1, r', \rho') + \psi(p, h, \rho) f(\rho, \rho') \\ \text{OPT}(k - 1, h - 1, r', \rho') + \psi(p, h, \rho) f(\rho, \rho') \end{cases}, \quad (9.9)$$

now has two cases on the right hand side. The first case represents expanding upon an optimal subproblem that has the same number of inflections, while the second case considers increasing the inflection count by one and builds upon optimal subproblems with fewer inflections. A subtle but important change is in the addition of the parameter k to the point adjacency set function $\mathcal{A}(k, h, r, \rho)$, along with a new constraint that is dependent on the value of k . If k is even then ρ must be strictly decreasing with respect to ρ' ,

$$\rho < \rho'$$

else it must be non-decreasing

$$\rho \geq \rho'$$

to be considered a valid adjacency. This forces ρ to be either strictly increasing or strictly non-decreasing until the inflection count changes, appropriately reversing the sign of the second derivative.

Examining subproblem traversal and feasibility, problems are examined in order of increasing h for increasing values of k . The first case of constant k is identical to the three-dimensional optimal subproblem and is likewise satisfiable. The second case is dependent on $\text{OPT}(k - 1, \cdot, \cdot, \cdot)$, which by the traversability order will always be computed and available when evaluating $\text{OPT}(k, \cdot, \cdot, \cdot)$.

9.4. Multi-view perceptual grouping and generatrix recovery evaluation

The simultaneous n -view perceptual grouping and generatrix recovery technique allows information from n -views to be aggregated into to a common three-dimensional generatrix

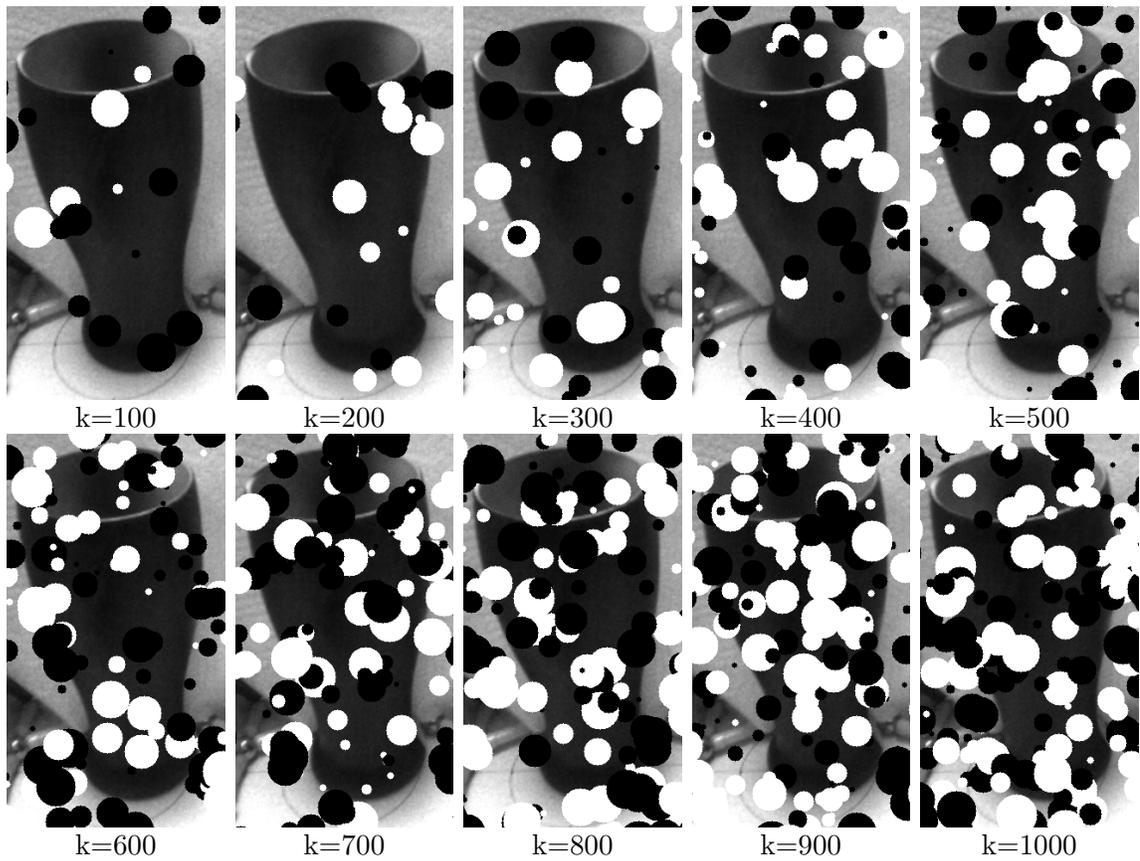


Figure 9.6: Speckle noise applied to a single view with varying densities. Random white and black speckle noise is added to a dataset view parameterized by speckle density k , the number of speckles per image. The 10 non-zero experiment speckle densities are visualized through application to a single dataset view.

Speckle density	Mean reconstruction error (mm) by number of images used									
	1	2	3	4	5	10	25	100	500	900
0	0.1	0.1	0.1	0.2	0.2	0.2	0.0	0.0	0.0	0.0
100	0.8	0.6	0.4	0.3	0.2	0.3	0.1	0.0	0.0	0.0
200	0.2	0.2	0.3	0.4	0.5	0.3	0.2	0.1	0.0	0.1
300	0.3	0.9	0.6	0.3	0.3	0.4	0.2	0.1	0.1	0.0
400	7.3	5.8	0.4	0.3	0.2	0.5	0.3	0.1	0.1	0.1
500	5.3	10.0	5.3	0.4	0.4	0.4	0.2	0.0	0.1	0.1
600	6.9	2.3	0.7	0.9	2.1	0.5	0.4	0.2	0.1	0.0
700	2.3	10.6	2.9	2.8	2.4	2.2	2.9	0.3	0.1	0.1
800	6.8	1.7	2.6	2.7	2.4	2.1	0.3	0.3	0.2	0.1
900	12.0	10.2	12.2	12.4	8.8	1.5	1.6	0.3	0.2	0.1
1000	14.6	13.8	3.3	1.7	1.6	4.0	0.7	0.3	0.3	0.2

Table 9.1: Mean SoR reconstruction error by speckle density and aggregated image count. SoR reconstruction error is measured as the difference between the groundtruth generatrix and the generatrix estimated using n images with speckle density k . Increasing speckle noise as parameterized by speckle density increases the mean SoR reconstruction error, while increases the number of images used for the reconstruction decreases this error.

space. This provides a mechanism to combine SoR structure evidence from many noisy images and extract a globally optimal generatrix, provided the poses of the SoR are known. The optimal generatrix is considered to be the one that passes through the most response energy, subject to smoothness constraints.

The performance of this generatrix recovery technique is evaluated with respect to two parameters, the number of images n used and the amount of image noise. The input images of prior work (Figure 1.6) allow for reliable segmentation and contour grouping, with high contours and crisp contours. To violate this assumption as definitively as possible, image noise is introduced that is modeled by random white and black speckles. It is parameterized by speckle density k , the number of speckles per image. The performance metric is the mean reconstruction error with respect to the groundtruth generatrix.

Ten non-zero speckle densities are applied to all 900 views of the view-plane dataset, as illustrated in Fig. 9.6. Figure 9.1 shows the resulting reconstruction errors for all speckle densities and select image counts. Reconstruction errors achieve values of less than 0.2mm in the limit with 900 images. With speckle density of 500 and less, 5 images is sufficient to achieve reconstruction errors of less than 0.5mm. Even with the highest density of 1000, less than 1.0mm error is achieved with 25 images, and less than 0.5mm with 100 images. Figure

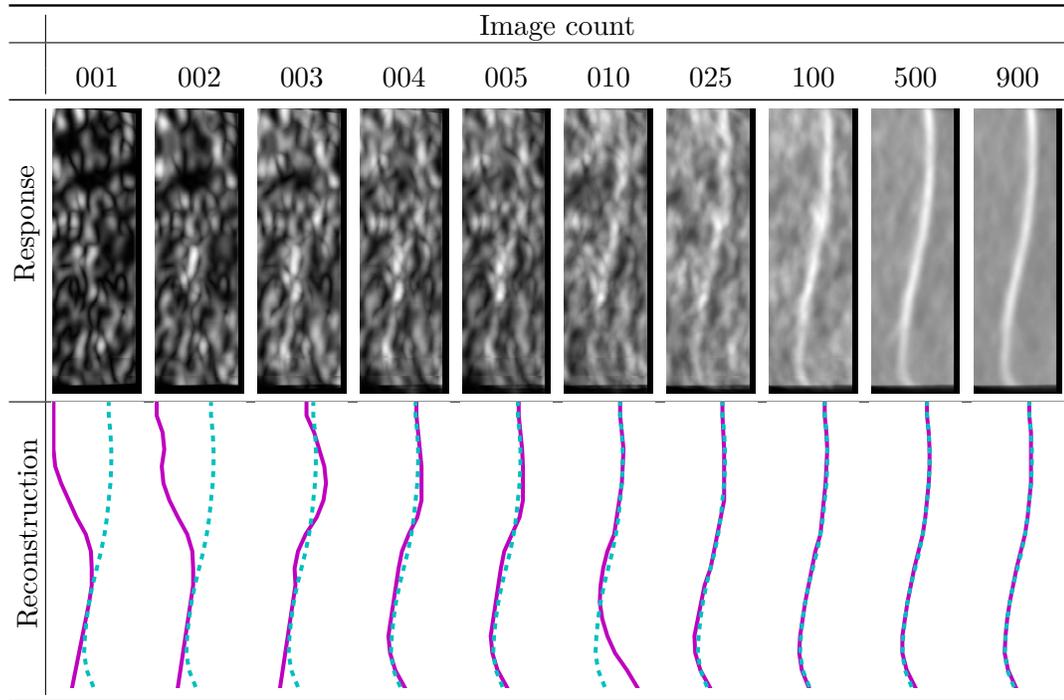


Figure 9.7: Aggregated SoR responses and reconstructions for varying input image counts. The SoR reconstruction procedure operates over an image response that is aggregated in three-dimensional generatrix space. The optimal generatrix is considered to be the one that passes through the most response energy, subject to smoothness constraints. As the number of images aggregated increases, the less sensitive the response to image noise and the better the reconstruction. The solid line represents the reconstructed generatrix as compared to the dotted line representing the groundtruth generatrix.

9.7 illustrates the incremental effect of adding images on both the aggregated response and the resulting SoR generatrix reconstruction.

9.5. Summary

This chapter presented a method for SoR apparent contour perceptual grouping and shape recovery that operates over multiple views with extreme speckle noise corruption and known SoR poses. This is a departure from prior shape recovery techniques, which are not designed for such noisy inputs, and are typically evaluated on high-accuracy contours that are either easily and automatically extracted, manually annotated, or entirely synthetic. A dynamic programming technique was presented that operates in a common 3D generatrix space, aggregates evidence from multiple images, and enforces global smoothness and shape complexity constraints. It was shown that impressive reconstruction results can be achieved with just a few noisy images, with results that converge very close to the groundtruth as more images are used.

10 Perceptual grouping and shape recovery of optically challenging SoRs

Parallax is an effect whereby the position of an object appears differently when viewed from different view points. Such a displacement is dependent on the distance of the object, with nearby objects exhibiting a larger apparent motion than those at a distance. Relating the parallax between corresponding image points to their depth is the basis of stereo depth perception. This chapter uses this relationship to compute a cue to the presence of an object protruding from a known supporting plane, expanding on the work presented in Phillips et al. (2011).

The parallax cue allows the supporting plane points to be discriminated from off-plane points by checking for photometric consistency between corresponding plane points across images. Any points that disobey the parallax mapping are likely to produce photometric inconsistency, providing a response even for edges and regions that are non-salient in a single image. This implies that the applicability of the cue is not limited to just opaque surfaces, but encompasses transparent and other non-Lambertian surfaces as well.

If the SoR is assumed to be upright with respect to the known supporting plane, the only remaining pose parameters are encoded in the unknown intersection point between the supporting plane and the SoR central axis. Constraining the intersection point to the supporting plane leaves 2-DoF in the SoR pose. The projection of an imaged SoR onto a $\hat{\mathbf{z}}$ -aligned unit cylinder, as defined in Section 4.6, yields a profile that is symmetric about a vertical line with constant azimuth in cylindrical coordinates. Reprojecting the parallax cue map in this manner for the two images and searching such a symmetric signal allows

the 3D-axis triangulation used in Section 6.2 to recover the remaining pose translation parameters. The SoR pose and the parallax cue used in conjunction with the techniques of Chapter 9 allows the generatrix recovery of visually challenging surfaces of revolution.

10.1. Parallax of a plane

The parallax displacement of a point between two images is dependent on the distance of the point from each camera. In the case of a plane in 3D, these point-camera distances and the corresponding parallax disparities are described by a concise linear form. Assuming a 3D plane p in known pose with respect to a camera a , the bijection that relates 2D image points with 2D points on the planes surface is described in Equation 3.8 as a 3×3 homography matrix ${}^p\mathbf{H}_a$. A homogeneous plane point ${}^p\mathbf{x}$ is mapped to the homogeneous image point ${}^a\mathbf{x}$ in camera a by

$$\mathbf{x}_a = {}^a\mathbf{H}_p {}^p\mathbf{X} \quad (10.1)$$

Combining this equation with the homography ${}^p\mathbf{H}_b$ from an image point in camera b to the plane produces the homography between the image of the plane in b to the image of the plane in a as

$${}^a\mathbf{x} = \underbrace{{}^a\mathbf{H}_p {}^p\mathbf{H}_b}_{{}^a\mathbf{H}_b} {}^b\mathbf{x}, \quad (10.2)$$

with the parallax displacement computed as

$${}^a\delta_b = {}^b\mathbf{x} - {}^a\mathbf{x} \quad (10.3)$$

Figure 10.1 illustrates how a 2D grid imposed upon the 3D plane maintains a grid structure as it is displaced between views via the parallax effect.

10.2. Planar parallax photometric consistency

Given the plane pose in both cameras a and b with images \mathbf{I}_a and \mathbf{I}_b , the plane point ${}^a\mathbf{x}$ in \mathbf{I}_a that corresponds to plane point ${}^b\mathbf{x}$ in \mathbf{I}_b is given in Equation 10.2 as homography ${}^a\mathbf{H}_b$,

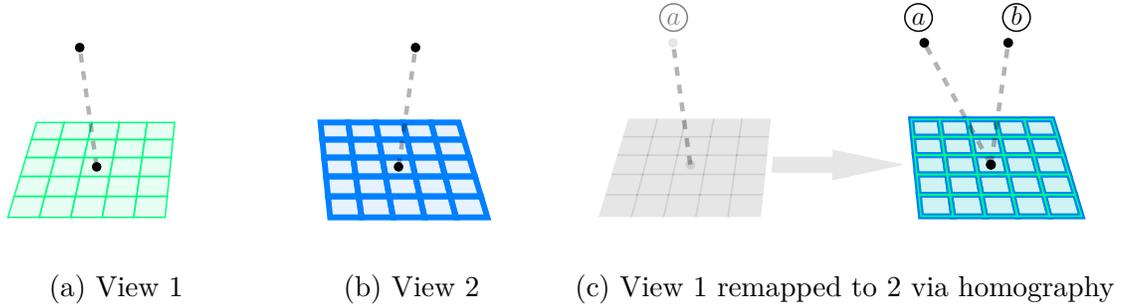


Figure 10.1: Geometric consistency and the parallax effect under supporting plane homography. The same plane, point and line are imaged by a camera that has translated to left between view 1 (a) and view (2). Remapping view 1 to view 2 (c) via the planar homography, the image of the plane from view 1 coincides perfectly with that of view 2, however the plane point displays strong parallax. The distance between point a and point b is the parallax displacement.

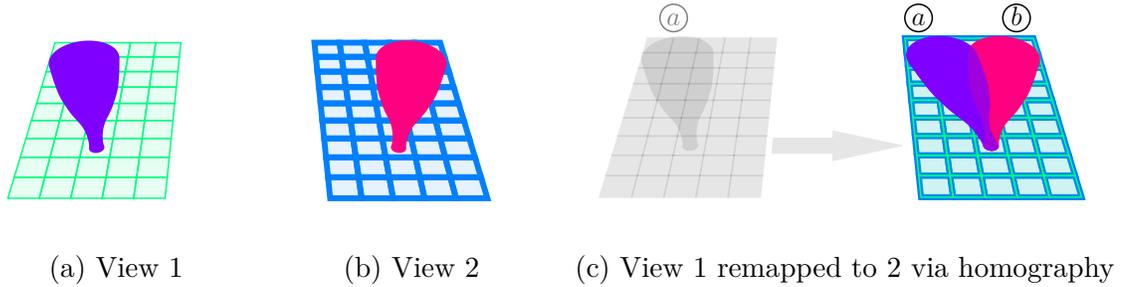


Figure 10.2: Photometric inconsistency of SoR profile under supporting plane homography. The same SoR and plane are imaged by a camera that has translated to the left between view 1 (a) and view (2). Remapping view 1 to view 2 (c) via the planar homography, the image of the plane from view 1 coincides perfectly with that of view 2, however the SoR profile displays strong parallax, and does not map onto itself.

called the *stereo inverse perspective map*. If plane image \mathbf{I}_b is remapped by ${}^a\mathbf{H}_b$, then the resulting image \mathbf{I}'_b should be identical to image \mathbf{I}_a , as illustrated in Figure 10.1. The two images are said to be *photometrically consistent* with respect to the homography.

10.3. Parallax cue computation

Figure 10.1 also shows that a line extending upward from the plane does not map onto itself under the supporting plane homography, but rather yields a photometric inconsistency between the image \mathbf{I}_a and the remapped image \mathbf{I}'_b . The presence of such an inconsistency has

been used previously as a reliable indicator of ground plane clutter in vehicular obstacle detection and avoidance (Storjohann et al., 1990; Mallot et al., 1991; Simond and Parent, 2007), but never as a segmentation energy map. Figure 10.2 shows the stereo inverse perspective map (SIPV) applied to a synthetic SoR mask computed for a stereo pair. Superimposing \mathbf{I}_a and the remapped \mathbf{I}'_b , as in Figure 10.1, illustrates how object points that do not belong to the supporting surface produce photometrically inconsistent boundaries and regions. The most notable are areas that contain binocular half-occlusions (Egnal and Wildes, 2002), where the supporting plane is seen in one view, but occluded by the object's surface in the other.

Representing the intensity of image \mathbf{I} at point \mathbf{x} as the function $\mathbf{I}[\mathbf{x}]$, the discrepancy image \mathbf{D}_a for image \mathbf{I}_a with respect to image \mathbf{I}_b is

$$\mathbf{D}_a [{}^a\mathbf{x}] = \left| \mathbf{I}_a [{}^a\mathbf{x}] - \mathbf{I}_b [{}^b\mathbf{H}_a {}^a\mathbf{x}] \right|. \quad (10.4)$$

The discrepancy image for \mathbf{I}_b is similarly defined as

$$\mathbf{D}_b [{}^b\mathbf{x}] = \left| \mathbf{I}_b [{}^b\mathbf{x}] - \mathbf{I}_a [{}^a\mathbf{H}_b {}^b\mathbf{x}] \right|, \quad (10.5)$$

which can equivalently be defined in terms of \mathbf{D}_a as

$$\mathbf{D}_b [{}^b\mathbf{x}] = \mathbf{D}_a [{}^a\mathbf{H}_b {}^b\mathbf{x}]. \quad (10.6)$$

These discrepancy images are used as the underlying energy map to compute the function ψ for the dynamic programming segmentation and generatrix recovery procedure of Chapter 9.

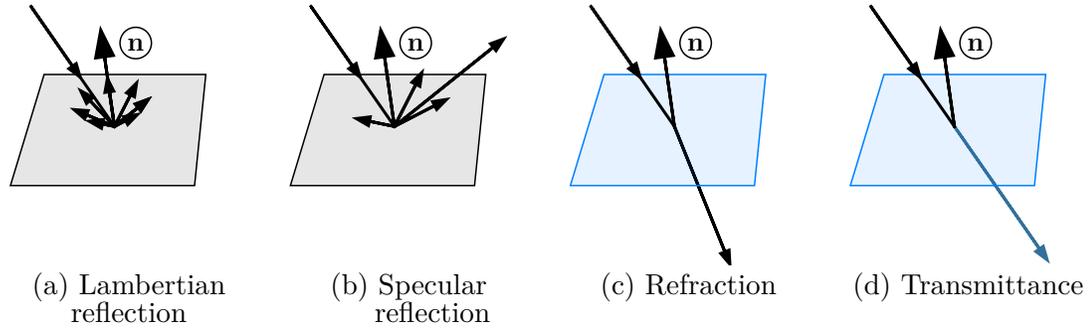


Figure 10.3: Lambertian vs. non-Lambertian surface properties. In a Lambertian reflection (a), the incoming ray is reflected isotropically outward in a sphere with equal intensities in all directions. In the case of specular reflections (b), rays of high intensity are reflected in some directions, but not others. Under refraction (c) a ray is bent as it passes through a medium. During the transmission of light (d), the intensity and color of the light may also change.

10.4. Non-Lambertian surface properties

One of the most prevalent assumptions in the development of vision algorithms and sensors is that of diffuse Lambertian surface reflectance, where light rays that strike an object's surface are isotropically reflected towards all viewpoints (see Figure 10.3a). This property allows a certain degree of photometric consistency when an object is viewed from different viewpoints. While the apparent shape of an imaged surface may change and some regions may become occluded, it is generally assumed that the imaged surface intensity values and gradients will be stable enough to model and match across views.

The strength of the parallax cue is that only the supporting plane is required to have the photometric consistency that is provided by a Lambertian surface. Photometric inconsistency in the foreground object can actually *increase* the cue response, as corresponding foreground object points match neither themselves nor the background. There are several ways in which a non-Lambertian surface can contribute to image intensity discrepancies between the two views. In the case of transparent objects, light is often reflected non-isotropically, leading to specularities arising from the fact that some viewpoints receive the reflected light at high intensity while others do not (see Figure 10.3b). Additionally, most

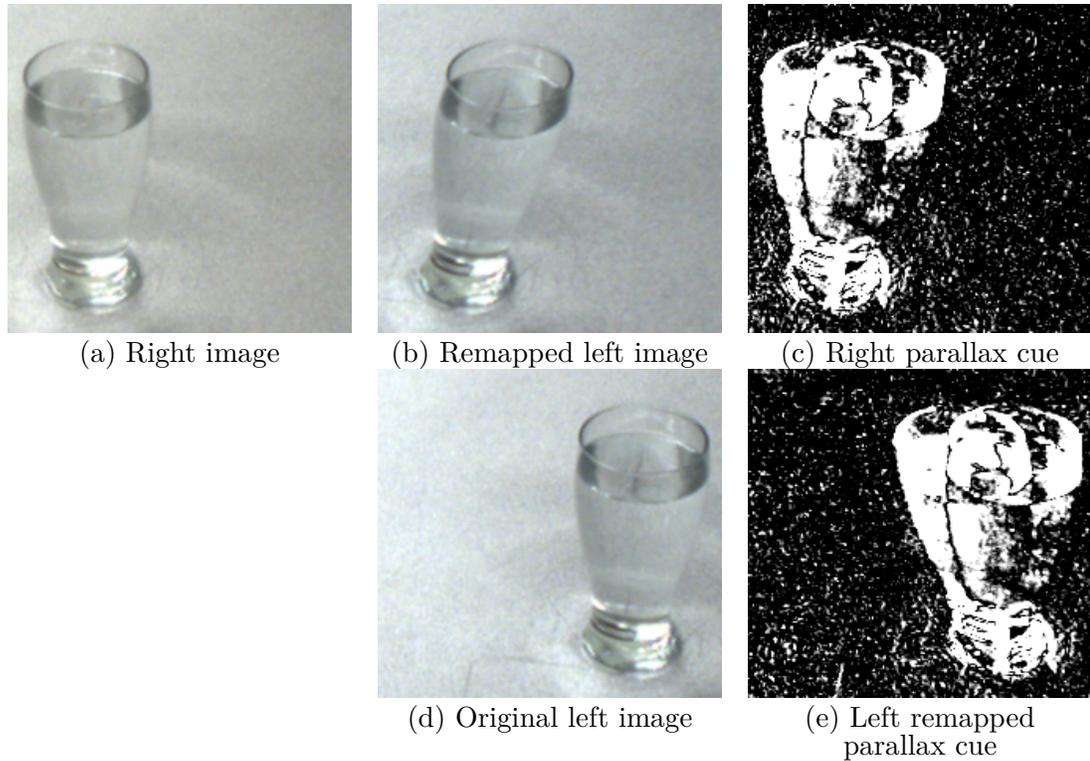


Figure 10.4: Planar homography warpings and the parallax cue. A planar homography exists between the supporting plane as viewed in the right (a) and left (d) images. The left image can be remapped (b) via this homography such that all plane points in the left image map to their corresponding points in the right image. As a result the photometric consistency of these corresponding plane points, the subtraction of the right image (a) from the remapped left image (b) yields high photometric inconsistency (c) for all points that are not a direct image of the supporting plane. This inconsistency is called the parallax cue, and can be computed in either the left or right image. If computed in the right image (c), it can be transferred to left image (e) by the inverse planar homography used to remap the left image to the right.

light rays are transmitted through the material and are bent (see Figure 10.3c) or diminished in intensity (see Figure 10.3d) based on the material's refractive index and transmittance.

Due to these properties, most light rays received at a specific viewpoint originate from the surfaces behind the object or from light sources in front of the object. In both cases, the light passing through the same transparent object surface point is likely to vary significantly with a change in viewpoint.

	Scene	Mean reconstruction error (mm) by number of images used									
		1	2	3	4	5	6	7	8	9	10
plain	black plain	2.3	1.4	1.1	1.0	1.8	1.7	1.7	1.8	1.5	1.3
	round	2.3	2.2	2.3	1.8	1.7	1.6	1.3	1.3	1.2	1.2
	clutter1	2.6	1.6	1.8	1.5	1.4	1.2	1.1	1.1	1.1	1.0
	clutter2	3.1	1.0	1.6	1.9	1.0	0.9	1.1	1.1	0.8	0.8
	clutter3	1.4	1.9	1.5	1.5	1.3	1.3	1.3	1.2	1.2	1.1
	clutter4	1.5	1.4	1.6	1.4	1.3	1.2	1.0	0.8	0.9	1.0
water	black	2.2	3.6	1.1	1.7	1.9	1.8	1.8	1.6	1.2	0.9
	round	1.9	2.1	1.7	1.6	1.6	1.5	1.4	1.3	1.3	1.2
	clutter1	3.3	1.8	3.3	1.3	1.5	1.4	1.5	1.5	1.4	1.4
	clutter2	2.9	1.8	2.0	1.7	1.5	1.3	1.0	1.0	0.9	0.9
	clutter3	1.3	0.8	0.9	0.8	1.1	0.7	0.6	0.7	0.6	0.7
	clutter4	1.1	1.0	1.3	1.0	1.1	1.0	1.0	0.8	1.0	1.0

Table 10.1: Mean SoR reconstruction error by scene and aggregated image count. SoR reconstruction error is measured as the difference between the groundtruth generatrix and the generatrix estimated using k images. While the various scene configurations produce different error baselines, increasing the number of images used for the reconstruction decreases the relative error for all scenes.

10.5. Transparent SoR perceptual grouping and generatrix recovery evaluation

Transparent object perception has received very sparse attention in the literature. Most approaches capture the geometric inconsistencies of glass objects in some manner, whether it be through the detection of sensor anomalies and failures (Klank et al., 2011; Lysenkov and Eruhimov, 2013; Lysenkov and Rabaud, 2013), explicit statistical modeling of highlights, caustics, and distortions, (McHenry et al., 2005; McHenry and Ponce, 2006; Kompella and Sturm, 2011) or more general adaptive learning techniques (Fritz et al., 2009; Wang et al., 2013). The parallax cue falls mostly in the “sensor anomaly” family of techniques, indirectly capturing light field distortions through the photometric inconsistency revealed by the planar homology.

The stereoscopic parallax cue provides a response map to use as inputs to the simultaneous n -view perceptual grouping and generatrix recovery technique presented in Chapter 9. The cue is generated for all images transparent SoR dataset (see Figure 10.4). The number of images used and the source scene configuration are the two parameters varied to evaluate the reconstruction technique’s performance using the parallax cue. There are 12 scene configurations with 10 images each, yielding a total of 120 samples.

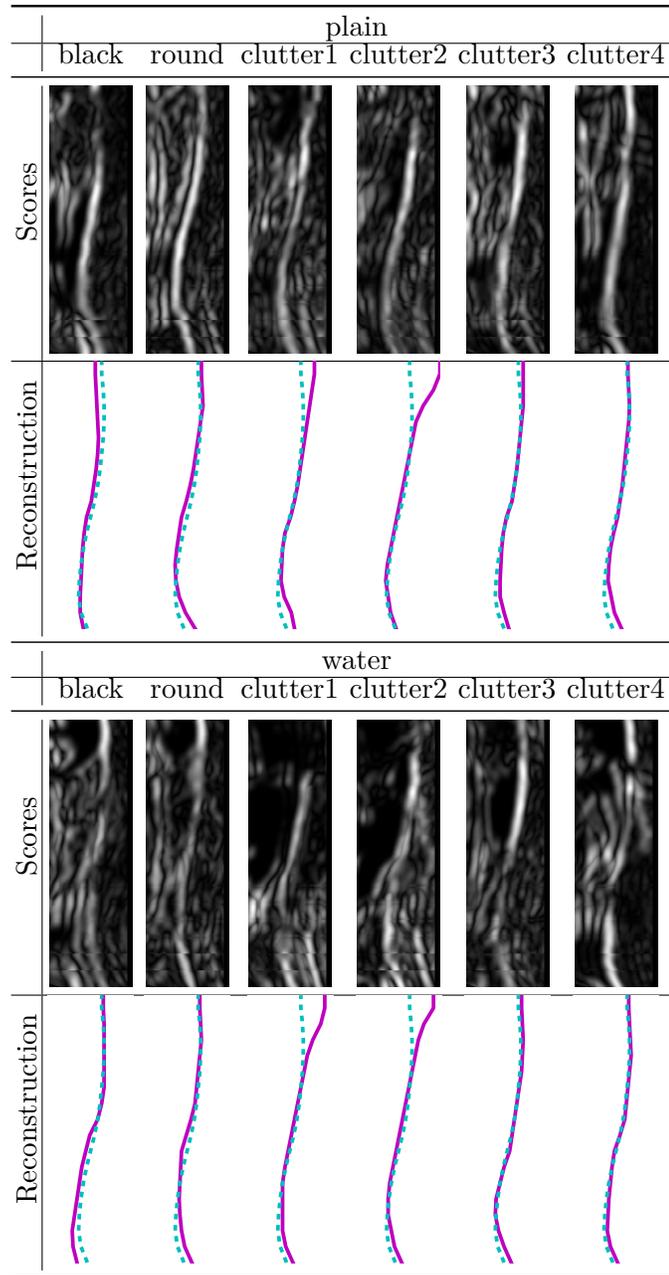


Figure 10.5: SoR responses and reconstructions for all scenes using a single input image. The SoR reconstruction procedure operates over an image response that is aggregated in three-dimensional generatrix space. The optimal generatrix is considered to be the one that passes through the most response energy, subject to smoothness constraints. While the major structure of the SoR is visible in most cases, there are many areas of signal dropouts and extreme noise, as reflected in the corresponding SoR reconstructions. The solid line represents the reconstructed generatrix as compared to the dotted line representing the groundtruth generatrix.

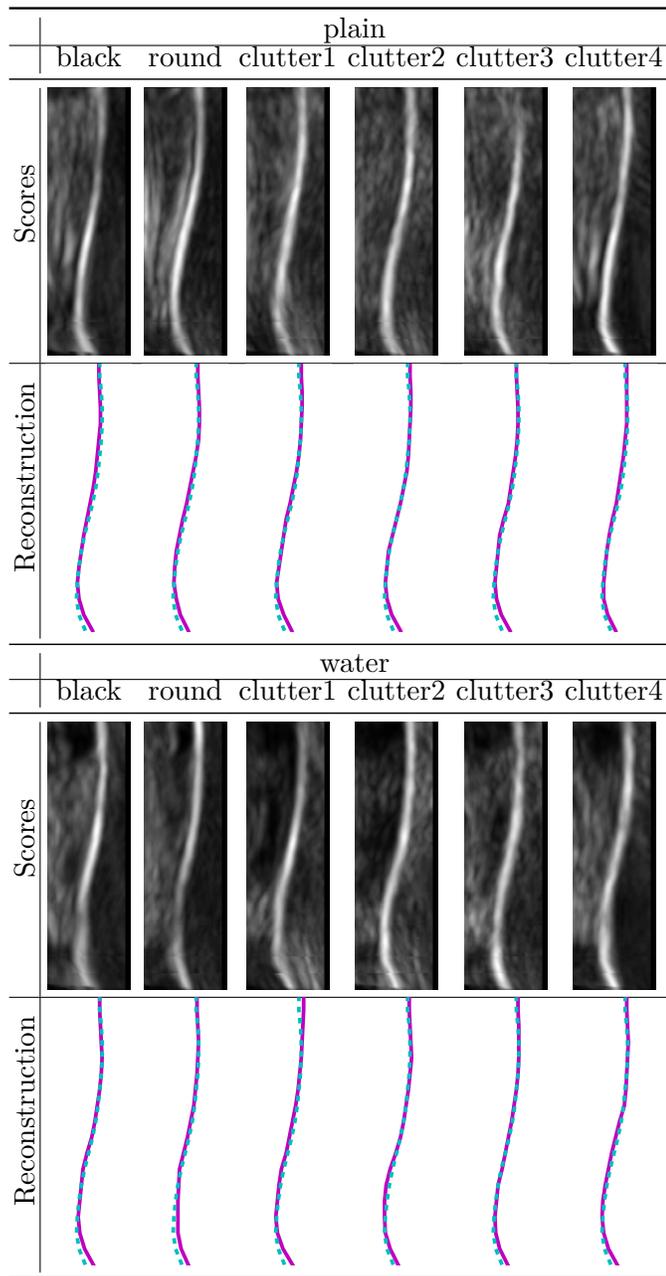


Figure 10.6: SoR responses and reconstructions for all scenes using a single input image. The SoR reconstruction procedure operates over an image response that is aggregated in three-dimensional generatrix space. The optimal generatrix is considered to be the one that passes through the most response energy, subject to smoothness constraints. Using 10 images per scene, the major structure of the SoR is salient over the signal noise contributed by each individual image. The corresponding SoR reconstructions (solid lines) show close fidelity to the groundtruth generatrix (dotted lines).

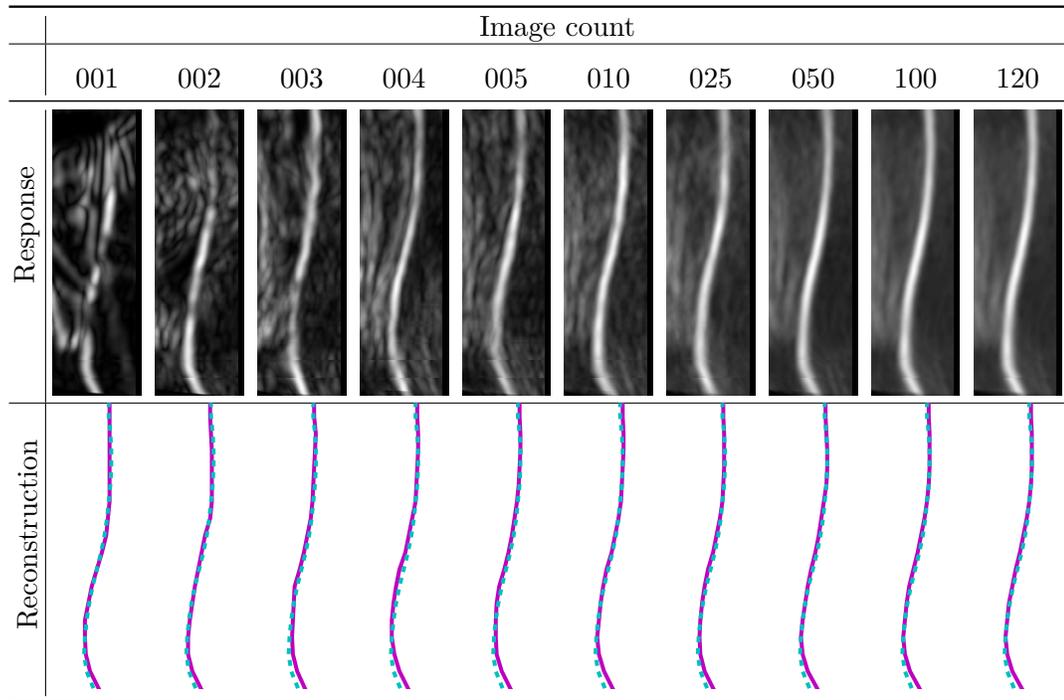


Figure 10.7: SoR parallax responses and reconstructions for varying image counts over all scenes. The SoR reconstruction procedure operates over an image response that is aggregated in three-dimensional generatrix space. The optimal generatrix is considered to be the one that passes through the most response energy, subject to smoothness constraints. As the number of images used is increased, the noise contribution of each individual image fades and the structure of the SoR becomes salient. The corresponding SoR reconstructions (solid lines) show increasing fidelity to the groundtruth generatrix (dotted lines).

Table 10.1 reports the reconstruction error by scene configuration and number of images used. Using all images, the average reconstruction error is under 1.5mm. For comparison, the groundtruth reconstructions over all 900 view-plane poses are typically within 0.4mm of average groundtruth reconstruction. Figure 10.5 shows the result of reconstructing each scene configuration with just one image, while Fig. 10.6 shows the result of using all 10 images per scene. Using all images for reconstruction (see Fig. 10.7), the reconstruction error drops below 1.0mm after 30 images, and ultimately converges to 0.9mm.

10.6. Summary

This chapter presented a parallax-based cue that increases the saliency of optically challenging SoRs, and showed that it can be used with the SoR apparent contour perceptual grouping and shape recovery technique of Chapter 9. Combining these methods, it was shown that impressive reconstruction results can be achieved for transparent SoRs with just a few images, with results that converge very close to the groundtruth as more images are used.

Part IV

Discussion and Conclusions

In summary, this thesis presented a Euclidean framework for SoR pose recovery, shape reconstruction and apparent contour perceptual grouping.

SoR forward-projection decomposition and profile–shape bijection

A pose-parameterized Euclidean SoR forward-projection decomposition (Chapter 4) was presented along with an SoR profile–shape bijection. These formulations provide a coherent and complete view of the SoR imaging and reconstruction process. They reveal additional geometric constraints that are exploited in single and multi-view contexts. Specifically, the explicitly parameterized forms were used to derive minimal problems that overcome the limitations of prior approaches with respect to SoR pose and shape recovery as well as perceptual grouping.

SoR pose and shape recovery

Methods for SoR pose and shape recovery were presented for three different image metrology tasks. None of the pose recovery techniques assume visible cross sections, special projective points, or a near-orthographic camera, as required in prior work. For all pose recovery tasks the 2D central-axis projection is recovered using a two-point minimal problem that exploits the fact that the projection of an SoR is symmetric on the image sphere. Task one requires two views with a known relative camera transform that is used to triangulate the 3D central axis. Task two recovers the pose of an SoR given its apparent contours and a known shape model using a one-point minimal problem. Task three relaxes the first task requirement of a known relative camera transform, using a structure-from-motion formulation to recover the SoR pose and shape. These methods were evaluated over a variety of noise conditions, poses, and baselines (for two-views), and were shown to provide stable and accurate results in the presence of moderate noise, with pose errors that generally increase with increasing camera-to-SoR distance and increasing baseline depth.

SoR perceptual grouping and shape recovery

Methods for SoR apparent contour perceptual grouping and shape recovery were presented for two metrology tasks. Both tasks have many SoR-views with known poses. The first task views are corrupted by extreme speckle noise, while the second task views contain SoRs with optically challenging transparent surfaces. Prior shape recovery techniques are not designed for such noisy inputs, and are typically evaluated on high-accuracy contours that are either easily and automatically extracted, manually annotated, or entirely synthetic.

A dynamic programming technique was presented that operates in a common 3D generatrix space, aggregates evidence from multiple images and enforces global smoothness and shape complexity constraints. This technique was evaluated on both the speckle-corrupted dataset and the transparent dataset after the application of a parallax-based cue to increase saliency. It was shown in both cases that impressive reconstruction results can be achieved with just a few noisy images, with results that converge very close to the groundtruth as more images are used.

10.6.1. Future work

The SoR perceptual grouping and shape recovery methods presented in this thesis currently require the full 5-DoF SoR pose in order to be applied. If the axis orientation of the SoRs is known, perhaps by an upright assumption with respect to a known supporting plane or gravity vector, then the only unknown pose parameter is the 3-DoF SoR position in space. By applying the cylindrical reprojection of Section 4.6 using the known axis orientation, all SoRs of this orientation become symmetric in the 2D cylindrical image about some vertical image line. Recovering the 2D central-axis projection is thus reduced to a 1-DoF search for the translation of this vertical line. Such a symmetric-axis search problem over unsegmented images has received a great deal of attention in the literature (Atallah, 1985; Reisfeld et al., 1995; Tsogkas and Kokkinos, 2012). Using the 2D central-axis projection recovered in this manner along with known relative camera poses, the SoR 3D central-axis

can be triangulated as in Chapter 6. In this way, it is possible to perform perceptual grouping and shape and pose recovery from noisy, unsegmented images.

Part V

Appendices

APPENDIX A : Contour generators and surfaces of revolution

A.1. Contour generator bijection

A.1.1. Contour generator by height

An apparent contour point ${}^o\mathbf{x} = (x, y, h)^\top$ is constrained by the tangent constraint,

$${}^o\mathbf{n}^\top({}^o\mathbf{x} - {}^o\mathbf{t}_c) = 0,$$

and the intersection constraint $x^2 + y^2 = r^2$.

Using the first equation, x is solved for as follows:

$$\begin{aligned} {}^o\mathbf{n}^\top({}^o\mathbf{x} - {}^o\mathbf{t}_c) &= 0 \\ {}^o\mathbf{n}^\top{}^o\mathbf{x} - {}^o\mathbf{n}^\top{}^o\mathbf{t}_c &= 0 \\ {}^o\mathbf{n}^\top{}^o\mathbf{x} &= {}^o\mathbf{n}^\top{}^o\mathbf{t}_c \\ (x, y, -r\rho)(x, y, h)^\top &= (x, y, -r\rho)(\hat{d}, 0, \hat{h})^\top \\ x^2 + y^2 - r\rho h &= x\hat{d} - r\rho\hat{h} \\ r^2 - r\rho h &= x\hat{d} - r\rho\hat{h} \\ x &= (r^2 - r\rho h + r\rho\hat{h})/\hat{d} \\ x &= (r^2 + r\rho(\hat{h} - h))/\hat{d}. \end{aligned}$$

With known x , the second equation is solved for y :

$$\begin{aligned} x^2 + y^2 &= r^2 \\ y^2 &= r^2 - x^2 \\ y &= \pm\sqrt{r^2 - x^2}. \end{aligned}$$

A.1.2. Normal recovery

The normal ${}^o\mathbf{n}$ projects to the canonical image plane as

$${}^o\mathbf{n}' = (0, n_y, n_z)^\top.$$

The component n_x is recovered as:

$$\begin{aligned}\mathbf{n}^\top \mathbf{v}_o &= 0 \\ (n_x, n_y, n_z)^\top (v_x, v_y, v_z) &= 0 \\ n_x v_x + n_y v_y + n_z v_z &= 0 \\ n_x v_x &= -(n_y v_y + n_z v_z) \\ n_x &= -(n_y v_y + n_z v_z) / v_x.\end{aligned}$$

A.1.3. Depth recovery

The 3D contour point ${}^o\mathbf{x}$ is constrained by:

$${}^o\mathbf{m}^\top {}^o\mathbf{x} = 0 \quad (\text{Tangent plane constraint})$$

and

$$\lambda {}^o\mathbf{v} + {}^o\mathbf{t}_c = {}^o\mathbf{x}. \quad (\text{Image ray constraint})$$

Solving the first equation for y yields:

$$\begin{aligned} {}^o\mathbf{m}^\top {}^o\mathbf{x} &= 0 \\ (-n_y, n_x, 0)^\top {}^o\mathbf{x} &= 0 \\ -n_y x + n_x y &= 0 \\ y &= n_y(n_x)^{-1}x. \end{aligned}$$

Solving for the second equation for λ yields:

$$\begin{aligned} \lambda {}^o\mathbf{v} + {}^o\mathbf{t}_c &= {}^o\mathbf{x} \\ \lambda \begin{pmatrix} v_x \\ v_y \\ v_z \end{pmatrix} + \begin{pmatrix} \hat{d} \\ 0 \\ \hat{h} \end{pmatrix} &= \begin{pmatrix} x \\ y \\ z \end{pmatrix} \\ \lambda v_y &= y \\ \lambda &= y(v_y)^{-1} \\ \lambda &= n_y(n_x v_y)^{-1}x \\ \lambda &= n_y(n_x v_y)^{-1}(\lambda v_x + \hat{d}) \\ \lambda &= \lambda v_x n_y(n_x v_y)^{-1} + n_y(n_x v_y)^{-1}\hat{d} \\ \lambda - \lambda v_x n_y(n_x v_y)^{-1} &= n_y(n_x v_y)^{-1}\hat{d} \\ \lambda(1 - v_x n_y(n_x v_y)^{-1}) &= n_y(n_x v_y)^{-1}\hat{d} \\ \lambda &= \frac{n_y(n_x v_y)^{-1}\hat{d}}{1 - v_x n_y(n_x v_y)^{-1}} \\ \lambda &= \frac{n_y \hat{d}}{n_x v_y - v_x n_y}. \end{aligned}$$

Substituting λ back into the image ray equation yields ${}^o\mathbf{x}$.

APPENDIX B : Symmetry rotation recovery and 3D axis triangulation

B.1. Symmetry refinement by use of infinitesimal rotations

B.1.1. Infinitesimal transforms

An infinitesimal rotation matrix is an antisymmetric matrix of the form

$$\Omega(\omega_x, \omega_y, \omega_z) = \begin{pmatrix} 0 & +\omega_z & -\omega_y \\ -\omega_z & 0 & -\omega_x \\ +\omega_y & +\omega_x & 0 \end{pmatrix}, \quad (\text{B.1})$$

that represents the differential change in a vector \mathbf{v} upon application of the corresponding infinitesimal transform

$$\mathbf{M} = \mathbf{I} + \Omega. \quad (\text{B.2})$$

In other words, if two vectors \mathbf{a} and \mathbf{b} are related by

$$\mathbf{a} = (\mathbf{I} + \Omega) \mathbf{b}, \quad (\text{B.3})$$

then

$$\mathbf{a} - \mathbf{b} = \Omega \mathbf{b} \quad (\text{B.4})$$

represents the differential change.

For small rotations, the infinitesimal transform $\mathbf{I} + \Omega$ approximates the behavior of a rotation matrix in $SO(3)$. This form is used to compute the values of the vector $\hat{\omega} = (\omega_x, \omega_y, \omega_z)$ that locally minimizes the rotation estimation residuals. The infinitesimal rotation vector $\hat{\omega}$ is then mapped back a rotation matrix \mathbf{R}_Δ in $SO(3)$ by Rodrigues' rotation formula.

B.1.2. Symmetric rotation infinitesimal form

The rotation from the representative pose to the symmetric pose encodes the roll and yaw of the camera. Accordingly, it is the composition of two rotations about the camera's z -axis and y -axis, with no rotation about the x -axis. This zero-degree rotation is incorporated into the infinitesimal rotation matrix of Equation B.1 by fixing $\omega_x = 0$.

Error in the estimate of the symmetric rotation $[\mathbf{R}_r]_k$ yields residual vectors between the aligned corresponding contours ${}^k\mathbf{A}$ and ${}^k\mathbf{B}$. Interpreting the contours as concatenated component vectors,

$${}^k\mathbf{A} = (\mathbf{x}_a, \mathbf{y}_a, \mathbf{z}_a)^\top \quad (\text{B.5})$$

$${}^k\mathbf{B} = (\mathbf{x}_b, \mathbf{y}_b, \mathbf{z}_b)^\top \quad (\text{B.6})$$

the residual vectors are expressed as

$$\mathbf{r} = (\mathbf{x}_a + \mathbf{x}_b, \mathbf{y}_a - \mathbf{y}_b, \mathbf{z}_a - \mathbf{z}_b)^\top. \quad (\text{B.7})$$

Considering these residual vectors to be the differential change due to the application of some refining infinitesimal transform $\mathbf{M}_\Delta = \mathbf{I} + \Omega$ that approximates \mathbf{R}_Δ , its corresponding infinitesimal rotation matrix Ω_Δ can be directly fit to the residual vectors as

$$\begin{pmatrix} \omega_y \\ \omega_z \end{pmatrix} = \begin{pmatrix} -\mathbf{z}_a & \mathbf{y}_a \\ \mathbf{0} & -\mathbf{x}_a \\ \mathbf{x}_a & \mathbf{0} \end{pmatrix}^+ \begin{pmatrix} \mathbf{x}_a + \mathbf{x}_b \\ \mathbf{y}_a - \mathbf{y}_b \\ \mathbf{z}_a - \mathbf{z}_b \end{pmatrix}, \quad \omega_x = 0, \quad (\text{B.8})$$

where operator $^+$ represents the pseudo-inverse.

The pseudo-inverse computes (ω_y, ω_z) to minimize the residual vectors, thereby minimizing the corresponding residual function Equation 6.7.

The derivation for the minimizing form of Equation B.8 is as follows

$$\begin{pmatrix} x_a + x_b \\ y_a - y_b \\ z_a - z_b \end{pmatrix} = \begin{pmatrix} 0 & +\omega_z & -\omega_y \\ -\omega_z & 0 & 0 \\ +\omega_y & 0 & 0 \end{pmatrix} \begin{pmatrix} x_a \\ y_a \\ z_a \end{pmatrix} \quad (\text{B.9})$$

$$\begin{pmatrix} x_a + x_b \\ y_a - y_b \\ z_a - z_b \end{pmatrix} = \begin{pmatrix} y_a\omega_z - z_a\omega_y \\ -x_a\omega_z \\ x_a\omega_y \end{pmatrix} \quad (\text{B.10})$$

$$\begin{pmatrix} x_a + x_b \\ y_a - y_b \\ z_a - z_b \end{pmatrix} = \begin{pmatrix} -z_a & y_a \\ 0 & -x_a \\ x_a & 0 \end{pmatrix} \begin{pmatrix} \omega_y \\ \omega_z \end{pmatrix} \quad (\text{B.11})$$

$$\begin{pmatrix} \mathbf{x}_a + \mathbf{x}_b \\ \mathbf{y}_a - \mathbf{y}_b \\ \mathbf{z}_a - \mathbf{z}_b \end{pmatrix} = \begin{pmatrix} -\mathbf{z}_a & \mathbf{y}_a \\ \mathbf{0} & -\mathbf{x}_a \\ \mathbf{x}_a & \mathbf{0} \end{pmatrix} \begin{pmatrix} \omega_y \\ \omega_z \end{pmatrix} \quad (\text{B.12})$$

$$\begin{pmatrix} \omega_y \\ \omega_z \end{pmatrix} = \begin{pmatrix} -\mathbf{z}_a & \mathbf{y}_a \\ \mathbf{0} & -\mathbf{x}_a \\ \mathbf{x}_a & \mathbf{0} \end{pmatrix}^+ \begin{pmatrix} \mathbf{x}_a + \mathbf{x}_b \\ \mathbf{y}_a - \mathbf{y}_b \\ \mathbf{z}_a - \mathbf{z}_b \end{pmatrix} \quad (\text{B.13})$$

The operator $^+$ represents the pseudo-inverse.

APPENDIX C : Two-point minimal correspondence problem for structure from
motion

C.1. Dual attitude recovery

The attitudes of two SoR views can be recovered from two contour point correspondences between the view a and b :

$$(\mathbf{x}_{a1}, \mathbf{l}_{a1}) \leftrightarrow (\mathbf{x}_{b1}, \mathbf{l}_{b1})$$

$$(\mathbf{x}_{a2}, \mathbf{l}_{a2}) \leftrightarrow (\mathbf{x}_{b2}, \mathbf{l}_{b2}).$$

Combining these correspondences with the two-view constraint of ρ from Equation 8.1 provides the system of equations,

$$\begin{aligned} k_{a1} \sin(\phi_a + \psi_{a1}) &= k_{b1} \sin(\phi_b + \psi_{b1}) \\ k_{a2} \sin(\phi_a + \psi_{a2}) &= k_{b2} \sin(\phi_b + \psi_{b2}), \end{aligned} \tag{C.1}$$

that allows us to solve for both ϕ_a and ϕ_b .

Applying the following change of variables:

$$\phi_a = \phi_{as} - \psi_{a1} \tag{C.2}$$

$$\phi_b = \phi_{bs} - \psi_{b1} \tag{C.3}$$

$$\psi_a = -\psi_{a1} + \psi_{a2} \tag{C.4}$$

$$\psi_b = -\psi_{b1} + \psi_{b2} \tag{C.5}$$

$$K_1 = \frac{K_{a1}}{K_{b1}} \tag{C.6}$$

$$K_2 = \frac{K_{a2}}{K_{b2}}, \tag{C.7}$$

and rearranging yields

$$K_1 \sin(\phi_{as}) = \sin(\delta_\phi + \phi_{as}) \quad (\text{C.8})$$

$$K_2 \sin(\phi_{as} + \psi_a) = \sin(\delta_\phi + \phi_{as} + \psi_b). \quad (\text{C.9})$$

Applying the following identities:

$$\sin(\delta_\phi + \phi_{as}) = \sqrt{-\cos^2(\delta_\phi + \phi_{as}) + 1} \quad (\text{C.10})$$

$$\sqrt{-K_1^2 \sin^2(\phi_{as}) + 1} = \cos(\delta_\phi + \phi_{as}) \quad (\text{C.11})$$

$$K_2 \sin(\phi_{as} + \psi_a) = \sin(\psi_b) \cos(\delta_\phi + \phi_{as}) + \sin(\delta_\phi + \phi_{as}) \cos(\psi_b), \quad (\text{C.12})$$

and solving in terms of ϕ_a yields

$$\cos_{\phi_{as}} b + \sin_{\phi_{as}} a = \sqrt{-\phi_{as}^2 c + 1}, \quad (\text{C.13})$$

with

$$a = -\frac{K_1 \cos(\psi_b)}{\sin(\psi_b)} + \frac{K_2 \cos(\psi_a)}{\sin(\psi_b)} \quad (\text{C.14})$$

$$b = \frac{K_2 \sin(\psi_a)}{\sin(\psi_b)} \quad (\text{C.15})$$

$$c = K_1^2. \quad (\text{C.16})$$

Rewriting into quadratic form,

$$\cos^2 \phi_{as} (b^2 - 1) + 2 \cos \phi_{as} \sin \phi_{as} ab + \sin^2 \phi_{as} (a^2 + c - 1) = 0, \quad (\text{C.17})$$

and substituting

$$x = \frac{\cos \phi_{as}}{\sin \phi_{as}}, \quad (\text{C.18})$$

yields the quadratic,

$$a^2 + 2abx + b^2x^2 + c - x^2 - 1 = 0, \quad (\text{C.19})$$

in terms of x .

This is solved using the quadratic formula, and then back-substituted to recover ϕ_a and ϕ_b .

C.1.1. Relative translation recovery

After ϕ_a and ϕ_b have been recovered, the apparent contour rays \mathbf{x}_a and \mathbf{x}_b in the SoR's coordinate system are known along with parameters μ_a and μ_b .

Using the contour generator parameterization of Equation 4.20,

$$\begin{pmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{pmatrix} = \begin{pmatrix} (\mu x + 1)\hat{d} \\ \mu y \hat{d} \\ \mu z \hat{d} + \hat{h} \end{pmatrix}, \quad \mu = f(\mathbf{x}, \mathbf{l}), \quad (\text{C.20})$$

substituting in the values,

$$(\hat{d}_a, 0, \hat{h}_a)^\top = (1, 0, 0)^\top \quad (\text{C.21})$$

$$(\hat{d}_b, 0, \hat{h}_b)^\top = (1 + \Delta_d, 0, \Delta_h)^\top \quad (\text{C.22})$$

and $\mathbf{x}_a, \mathbf{x}_b$ yields the two contour generator equations

$$\begin{pmatrix} \tilde{x}_a \\ \tilde{y}_a \\ \tilde{z}_a \end{pmatrix} = \begin{pmatrix} (\mu_a x_a + 1) \\ \mu_a y_a \\ \mu_a z_a \end{pmatrix}, \quad \begin{pmatrix} \tilde{x}_b \\ \tilde{y}_b \\ \tilde{z}_b \end{pmatrix} = \begin{pmatrix} (\mu_b x_b + 1)(1 + \Delta_d) \\ \mu_b y_b (1 + \Delta_d) \\ \mu_b z_b (1 + \Delta_d) + \Delta_h \end{pmatrix}. \quad (\text{C.23})$$

We want the values of Δ_d and Δ_h such that contour generator points describe the same

generatrix point as

$$(r_a, h_a) = (r_b, h_b). \quad (\text{C.24})$$

Starting with the radius equations for a and b , substituting in the contour generator equations, and solving for Δ_d yields the following:

$$r_a^2 = r_b^2 \quad (\text{C.25})$$

$$\tilde{x}_a^2 + \tilde{y}_a^2 = \tilde{x}_b^2 + \tilde{y}_b^2 \quad (\text{C.26})$$

$$\tilde{x}_a^2 + \tilde{y}_a^2 = ((\mu_b x_b + 1)^2 + (\mu_b y_b)^2)(1 + \Delta_d)^2 \quad (\text{C.27})$$

$$\frac{\tilde{x}_a^2 + \tilde{y}_a^2}{(\mu_b x_b + 1)^2 + (\mu_b y_b)^2} = (1 + \Delta_d)^2 \quad (\text{C.28})$$

$$\sqrt{\frac{\tilde{x}_a^2 + \tilde{y}_a^2}{(\mu_b x_b + 1)^2 + (\mu_b y_b)^2}} = 1 + \Delta_d \quad (\text{C.29})$$

$$\Delta_d = \sqrt{\frac{\tilde{x}_a^2 + \tilde{y}_a^2}{(\mu_b x_b + 1)^2 + (\mu_b y_b)^2}} - 1. \quad (\text{C.30})$$

Starting with the height equations and solving for a and b and solving for Δ_h yields:

$$\tilde{z}_a = \tilde{z}_b \quad (\text{C.31})$$

$$\tilde{z}_a = \mu_b z_b (1 + \Delta_d) + \Delta_h \quad (\text{C.32})$$

$$\Delta_h = \tilde{z}_a - \mu_b z_b (1 + \Delta_d). \quad (\text{C.33})$$

The translation is therefore recovered as

$$\Delta_d = \sqrt{\frac{\tilde{x}_a^2 + \tilde{y}_a^2}{(\mu_b x_b + 1)^2 + (\mu_b y_b)^2}} - 1. \quad (\text{C.34})$$

$$\Delta_h = \tilde{z}_a - \mu_b z_b (1 + \Delta_d). \quad (\text{C.35})$$

Bibliography

- M. J. Atallah. On symmetry detection. *IEEE Transactions on Computers*, C-34(7):663–666, July 1985.
- R. Bellman. The theory of dynamic programming. *Bulletin of the American Mathematical Society*, 60(6):503–515, Nov. 1954.
- G. M. Chaikin. An algorithm for high-speed curve generation. *Computer Graphics and Image Processing*, 3(4):346–349, 1974.
- R. Cipolla. *Active Visual Inference of Surface Shape*. PhD thesis, 1991.
- R. Cipolla and A. Blake. Surface shape from the deformation of apparent contours. *International Journal of Computer Vision*, 9(2):83–112, Nov. 1992.
- C. Colombo, A. Del Bimbo, and F. Pernici. Shape reconstruction from a single photograph for 3D object retrieval and visualization. In *Proceedings of the International Conference on Multimedia and Expo (ICME)*, volume 1, pages 897–9001, 2002.
- C. Colombo, D. Comanducci, A. Del Bimbo, and F. Pernici. Accurate automatic localization of surfaces of revolution for self-calibration and metric reconstruction. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, pages 55–55, June 2004.
- C. Colombo, A. Del Bimbo, and F. Pernici. Metric 3D reconstruction and texture acquisition of surfaces of revolution from a single uncalibrated view. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 27(1):99–114, Jan. 2005.
- C. Colombo, D. Com, and A. D. Bimbo. Camera calibration with two arbitrary coaxial circles. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 265–276, 2006.
- A. Derevianko, Y. Kuzmin, G. Burr, A. Jull, and J. Kim. AMS 14C age of the earliest pottery from the Russian Far East: 1996–2002 results. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 223–224: 735–739, 2004. Proceedings of the Ninth International Conference on Accelerator Mass Spectrometry.
- M. Dhome, J.-T. Lapresté, G. Rives, and M. Richetin. Spatial localization of modelled objects of revolution in monocular perspective vision. In *Proceedings of the European Conference on Computer Vision (ECCV)*, ECCV '90, pages 475–485, 1990.
- G. Egnal and R. Wildes. Detecting binocular half-occlusions: Empirical comparisons of five approaches. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 24(8):1127–1133, 2002.

- M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM (CACM)*, 24(6):381–395, June 1981.
- D. A. Forsyth, J. L. Mundy, A. Zisserman, and C. A. Rothwell. Recognising rotationally symmetric surfaces from their outlines. In G. Sandini, editor, *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 639–647, 1992.
- W. T. Freeman and E. H. Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 13(9):891–906, Sept. 1991.
- M. Fritz, M. Black, G. Bradski, S. Karayev, and T. Darrell. An additive latent feature model for transparent object recognition. In *Neural Information Processing Systems (NIPS)*, pages 558–566, 2009.
- P. Giblin and R. Weiss. Reconstruction of surfaces from profiles. Technical report, 1987.
- R. Glachet, M. Dhome, and J. T. Lapresté. Finding the perspective projection of an axis of revolution. *Pattern Recognition Letters*, 12(11):693–700, Nov. 1991.
- R. Glachet, M. Dhome, and J. T. Lapreste. *Computer Vision — ECCV’92: Second European Conference on Computer Vision Santa Margherita Ligure, Italy, May 19–22, 1992 Proceedings*, chapter Finding the pose of an object of revolution, pages 681–686. 1992.
- A. Gross and T. Boult. Recovery of shgcs from a single intensity view. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 18(2):161–180, Feb. 1996.
- J. R. Hurley and R. B. Cattell. The procrustes program: Producing direct rotation to test a hypothesized factor structure. *Behavioral Science*, 7(2):258–262, 1962.
- U. Klank, D. Carton, and M. Beetz. Transparent object detection and reconstruction on a mobile platform. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2011.
- V. Kompella and P. Sturm. Detection and avoidance of semi-transparent obstacles using a collective-reward based approach. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2011.
- I. Lysenkov and V. Eruhimov. Pose refinement of transparent rigid objects with a stereo camera. In M. Gavrilova, C. Tan, and A. Konushin, editors, *Transactions on Computational Science XIX*, volume 7870 of *Lecture Notes in Computer Science*, pages 143–157. 2013.
- I. Lysenkov and V. Rabaud. Pose estimation of rigid transparent objects in transparent clutter. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 162–169, May 2013.

- H. Mallot, H. Bulthoff, J. Little, and S. Bohrer. Inverse perspective mapping simplifies optical flow computation and obstacle detection. *Biological Cybernetics*, 64:177–185, 1991.
- K. McHenry and J. Ponce. A geodesic active contour framework for finding glass. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1038–1044, 2006.
- K. McHenry, J. Ponce, and D. Forsyth. Finding glass. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 973–979, 2005.
- P. R. S. Mendonca, K. Y. K. Wong, and R. Cippolla. Epipolar geometry from profiles under circular motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 23(6):604–616, June 2001.
- C. Phillips and K. Daniilidis. Absolute pose and structure from motion for surfaces of revolution: minimal problems using apparent contours. In *3D Vision (3DV), 2016 International Conference on*, pages 648–656, Oct. 2016.
- C. Phillips, K. Derpanis, and K. Daniilidis. A novel stereoscopic cue for figure-ground segregation of semi-transparent objects. In *IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 1100–1107, 2011.
- C. Phillips, M. Lecce, C. Davis, and K. Daniilidis. Grasping surfaces of revolution: Simultaneous pose and shape recovery from two views. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 1352–1359, May 2015.
- C. Phillips, M. Lecce, and K. Daniilidis. Seeing glassware: from edge detection to pose estimation and shape recovery. In *Proceedings of Robotics: Science and Systems*, 2016.
- J. Ponce, D. Chelberg, and W. Mann. Invariant properties of straight homogeneous generalized cylinders and their contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 11(9):951–966, Sept. 1989.
- D. Reisfeld, H. Wolfson, and Y. Yeshurun. Context-free attentional operators: The generalized symmetry transform. *International Journal of Computer Vision*, 14(2):119–130, 1995.
- H. Sato and T. Binford. On finding the ends of straight homogeneous generalized cylinders. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 695–698, June 1992.
- H. Sato and T. O. Binford. Finding and recovering shgc objects in an edge image. *CVGIP: Graphical Models and Image Processing*, 57(3):346–358, May 1993.
- N. Simond and M. Parent. Obstacle detection from IPm and super-homography. In *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 4283–4288, Oct. 2007.

- K. Storjohann, T. Zielke, H. A. Mallot, and W. von Seelen. Visual obstacle detection for automatically guided vehicles. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 761–7662, May 1990.
- S. Tsogkas and I. Kokkinos. *Learning-Based Symmetry Detection in Natural Images*, pages 41–54. 2012.
- S. Utcke and A. Zisserman. Projective reconstruction of surfaces of revolution. In B. Michaelis and G. Krell, editors, *Pattern Recognition*, pages 265–272, 2003.
- T. Wang, X. He, and N. Barnes. Glass object segmentation by label transfer on joint depth and appearance manifolds. *Proceedings of the International Conference on Image Processing (ICIP)*, pages 2944–2948, Sept. 2013.
- K.-Y. Wong, P. Mendonca, and R. Cipolla. Camera calibration from surfaces of revolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 25(2):147–161, Feb. 2003.
- K.-Y. K. Wong, P. R. S. Mendonça, and R. Cipolla. Reconstruction of surfaces of revolution from single uncalibrated views. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 93–102, 2002.
- A. Zisserman, J. Mundy, D. Forsyth, J. Liu, N. Pillow, C. Rothwell, and S. Utcke. Class-based grouping in perspective images. In *IEEE International Conference on Computer Vision (ICCV)*, pages 183–188, June 1995.