Shape Recovery by Exploiting Planar Topology in 3D Projective Space

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By
Po-Lun Lai

Graduate Program in Geodetic Science

The Ohio State University
2010

Dissertation Committee:
Alper Yilmaz, Advisor
Tony Schenk
Rongxing Li
ABSTRACT

In the fields of photogrammetry and computer vision, three-dimensional (3D) shape recovery has remained an active research topic over the past decades. Accompanied by the boost in the development of sensor technology, considerable efforts have been expanded to image-based shape recovery. However, most of the approaches still rely on calibrated cameras with known orientations to establish the transformation between the object space and image space, keeping them from practical when the camera interior and exterior parameters are unavailable.

In this research, a novel approach is developed for the recovery of 3D object shape using uncalibrated multiple-view images. The approach is based on the assumption that the 3D projective space is composed of 2D discrete projective subspaces. In the designed framework, a 2D subspace corresponds to a set of hypothetical planes which creates cross-sections by slicing the objects in the scene. The images of such cross-sections are obtained by planar projective transforms which are estimated from the points defined on these hypothetical planes. A stack of these cross-sections provides a projective recovery of the 3D object shape. The resulting recovery becomes an affine or metric recovery when stack of cross-sections are transformed to an affine or ortho-rectified image respectively, or when absolute ground information is provided.

In this dissertation, two possible formations of subspaces are proposed, and several experiments using image sets of different characteristics are conducted to evaluate
the performance of the proposed method. Generated 3D shapes, when qualitatively examined or quantitatively compared against ground-truth, show promising recovery performance.
This dissertation is dedicated to my wife Wan-Chen, and our families in Taiwan...
ACKNOWLEDGMENTS

I would like to give my sincere appreciation to my advisor, Dr. Alper Yilmaz, who has profoundly changed my life. He guided my study at OSU with his unreserved patience and encouragement since 2007. He also set up a sample for me to follow with his excellence and diligence to explore new research ideas.

I would also like to thank Dr. Tony Schenk, Dr. Rongxing Li and Dr. Morton O’Kelly for serving on my dissertation committee as reviewers and examiners. Their valuable comments and suggestions not only improve and complete my dissertation, but also point out a direction to my future research.

Special thanks are given to the current and previous members of the Photogrammetric Computer Vision Lab. They never hesitate to share their experiences and innovations, and have promoted a unique atmosphere of academic excellence. In addition, their kindness and friendship have made the lab a warm place to work and stay.

My appreciation also goes to the Ministry of Education, Taiwan, that had partially funded my expenses during 2006 - 2008 with the studying abroad scholarship program (SAS-95214-1-US-22).

I dedicate this achievement to my parents and my wife for all the love, encouragement, and inspiration they have provided during my pursuit of knowledge. I am, and will always be, grateful for their unconditional support.
VITA

January 8, 1973 ......................... Born - Pingtung, Taiwan

June 1997 ............................... B.S., Civil Engineering  
                                      National Taiwan University 

June 1999 ............................... M.S., Civil Engineering  
                                      National Taiwan University 

June 2008 ............................... M.S., Geodetic Science  
                                      The Ohio State University. 

September 2004 to present  ............. Graduate Research Associate, Department of CEEGS, The Ohio State University. 

PUBLICATIONS

Research Publications


FIELDS OF STUDY

Major Field: Geodetic Science

Studies in Photogrammetric Computer Vision: Prof. Alper Yilmaz
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Abstract</th>
<th>ii</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>v</td>
</tr>
<tr>
<td>Vita</td>
<td>vi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>x</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xi</td>
</tr>
</tbody>
</table>

Chapters:

1. Introduction ........................................ 1  
   1.1 Three dimensional recovery from single view .......... 4  
   1.2 Multi-view approaches ................................ 5  
       1.2.1 Structure from motion .......................... 6  
       1.2.2 Plane sweep stereo ............................. 8  
       1.2.3 Voxel coloring and space carving ................. 9  
       1.2.4 Shape from silhouettes ......................... 10  
   1.3 Proposed approach .................................. 12  
   1.4 Overview .......................................... 14  

2. Fundamentals of plane projective geometry ............... 16  
   2.1 Homogeneous Coordinates ............................ 16  
   2.2 Projection from 3D object space to 2D image .......... 18  
   2.3 Vanishing Point and Vanishing Line .................. 20  
       2.3.1 Vanishing Point ............................... 21  

viii
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td>82</td>
</tr>
<tr>
<td>Quality indices of the shape recovery.</td>
<td>82</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Pinhole camera geometry.</td>
<td>19</td>
</tr>
<tr>
<td>2.2 Vanishing points and vanishing line.</td>
<td>21</td>
</tr>
<tr>
<td>2.3 Homography between two images induced by a scene plane $\pi$.</td>
<td>24</td>
</tr>
<tr>
<td>3.1 Subspace topology for the parallel planes.</td>
<td>30</td>
</tr>
<tr>
<td>3.2 Subspace topology for the rotated planes.</td>
<td>31</td>
</tr>
<tr>
<td>3.3 Geometry of parallelogram and vanishing points.</td>
<td>34</td>
</tr>
<tr>
<td>3.4 Geometry of parallelogram and vanishing points.</td>
<td>36</td>
</tr>
<tr>
<td>3.5 Diagram for the scale ratio in Lemma 2.</td>
<td>38</td>
</tr>
<tr>
<td>3.6 The geometric relationship between the vanishing point of a reference direction and the vanishing line of the planes orthogonal to the direction.</td>
<td>39</td>
</tr>
<tr>
<td>3.7 False estimation of vanishing line for cropped image.</td>
<td>41</td>
</tr>
<tr>
<td>3.8 Geometry of the pixels under rotated plane topology.</td>
<td>44</td>
</tr>
<tr>
<td>4.1 Generation of an object cross-section.</td>
<td>48</td>
</tr>
<tr>
<td>4.2 The subspaces composed of planes generated from (a) translating and (b) rotating a reference plane.</td>
<td>51</td>
</tr>
<tr>
<td>4.3 Bounding-curves with the subspaces of (a) parallel and (b) rotated planes.</td>
<td>52</td>
</tr>
</tbody>
</table>
4.4 Error between the estimated and the true vanishing point. ....... 56
4.5 The exact projection of the synthetic points. ......................... 58
4.6 Errors on the scale factors under varying angle (α) of the error vector when \( r = 500 \). ....................................................... 59
4.7 The errors on the estimated pixels increase with \( Z \). .......... 59
4.8 The distances from the estimated image pixels to the actual pixels at plane \( Z = 0.2 \) for different \( r \) and \( α \) values. .................. 60
4.9 (a) The grid pixels transformed from another image using homography estimated from erroneous pixels. (b) Projection of the pixels in (a) onto the image of the ground plane. .......................... 61
4.10 Exact projection of the synthetic points. .............................. 62
4.11 Errors on the scale factors for the rotated planes with a change in the angle (α) of the error vector for \( r = 500 \). ................... 63
4.12 The errors on the estimated pixels increase with \( θ \). .......... 64
4.13 The distances from the estimated image pixels to the actual pixels at plane \( θ = \pi/4 \) under different \( r \) and \( α \). ...................... 65
4.14 (a) The grid pixels transformed from another image using homography estimated from erroneous pixels. (b) Projection of the pixels in (a) onto the image of the ground plane. .......................... 66
4.15 Work flow of the proposed approach. ............................... 67
5.1 Shape recovery for a toy. ............................................. 69
5.2 Test images for human shape recovery using parallel planes. .... 71
5.3 Recovered shape for a human body using parallel planes. ........ 72
5.4 Shape recovery for human body using rotated planes. .......... 73
5.5 Recovered shape for a human body using rotated planes. 74
5.6 Two images of Taj Mahal, India, selected and downloaded from the search result provided by Google©. 75
5.7 Recovered shape of Taj Mahal 76
5.8 Images of a building. 78
5.9 Recovered shape of the building. 79
5.10 Elevation grids created from LiDAR data and the recovered 3D model. 80
5.11 Two images from the temple dataset provided on [63]. 81
5.12 Reconstructed temple models. (a) The dense 3D point cloud generated from the proposed approach. (b) 3D model from laser scanning provided by [63]. 82
CHAPTER 1

Introduction

Recovering 3D world scenes from multiple images has long been a popular topic and received a great research interest in the past decades. Driven by the increasing mapping demand by the public, numerous military and civilian applications, and the evolution in technologies have generated recent thrusts to 3D recovery research. For instance, fast recovery of 3D scene or object shape can now be directly achieved by 3D scanners. However, the equipments are usually costly. Alternatively, a main line of research is focusing on recovering 3D shapes by analyzing image sequences, taking the advantages that cameras are cheaper to obtain and are lightweight. In addition, cameras nowadays are precise, providing enough resolution for most general applications. As a result, numerous research efforts have been dedicated to image-based modeling for creating 3D models from one or more images, and have found their applications including, but not limited to object tracking [42], scene understanding [13], shadow removal [72] [32], 3D scene recovery [53], action recognition [82] and scene modeling [10] [70] [54].

An ideal modeling system that could provide a general solution to the shape recovery is expected to satisfy several criteria including [47]:


1
1. **Automatic:** The automation limit the reconstruction time and expert technical knowledge involved.

2. **Accurate:** The recovery should correctly approximate the scene geometry and meet expected accuracy.

3. **Reliable:** Under situations where distinct features or surface texture are absent, or changes in appearance due to illumination and surface properties are present, the recovery remains executable.

4. **Photo-realistic:** The recovery should synthesis an equivalent visual quality to a real world scene.

5. **Versatile:** The recovery should work on multiple scale scenes and objects.

6. **Computationally Efficient:** A real-time system is always preferred.

Though these characteristics are not currently all realized by a single system, recent image-based approaches using fast and high resolution cameras at affordable rates have proven to meet most of the criteria and generated satisfactory results. Different approaches based on different sensors and assumptions are currently adopted according to the practical application priorities.

A popular approach in 3D scene recovery from images is to estimate the relative camera poses and calibrations prior to computing the scene structure. Traditionally, cameras are calibrated by exploiting measured features such as the ones on a calibrating box or a planar target array using a bundle adjustment framework [57]. Alternatively, calibration can be performed by the geometry formed from vanishing
points in orthogonal directions constructing a calibrating conic [28]. Recently, Grammatikopoulos et al. [24] proposed to use two vanishing points to calibrate cameras. Similar to calibration, camera pose can be estimated from point correspondences using a bundle adjustment framework which iteratively adjusts an initial camera pose and minimizes a non-linear cost [45]. For pose estimation, Ansar and Daniilidis proposed a linear approach utilizing both point and line features [2]. Other camera calibration and pose estimation techniques include factorizing the homography matrix [76] [49] and exploiting the epipolar geometry computed from dynamic silhouettes [69] under circular motion [30].

Traditionally, recovering 3D scene from cameras with known poses and calibrations is achieved by triangulation of point correspondences [29] [78] [12]. Specifically, triangulation refers to determining the 3D position of two corresponding pixels by intersecting the rays back-projected from images using the camera projection matrices. This operation requires sub-pixel localization of image pixels as well as precise estimation of camera parameters. Methods exploiting triangulation usually suffer from the sparseness of pixel correspondences, hence cannot provide dense 3D recovery. In addition, lack of a mechanism to handle occlusions may result in abrupt shapes that do not conform to the objects in the scene.

Alternative to triangulation, 3D scene can also be recovered by a variety of methods. Categorizing these approaches into some specific classes can be a difficult task, since many of these approaches are combinations of several techniques and properties. For instance, an approach may use a calibrated or an uncalibrated image, and present the object or scene by means of a depth map or surface models. In the following sections, related works under some named categories for object or scene recovery with
their strength and limitations are discussed. The approaches mentioned, however, is far from a complete list. For more details on comparison and evaluation of the state-of-the-art methods, the reader is referred to the comprehensive surveys provided in [74], [18], [64] and [47].

1.1 Three dimensional recovery from single view

Given only one view, a scene can be reconstructed from identifiable geometric entities such as lines, planes and spheres by computing their spatial layout. Inferences about the 3D structure can also be made from a single uncalibrated camera if vanishing points and lines can be computed. Generally, most approaches assume a simple camera model with zero skew and aspect ratio of one, and the internal parameters are estimated using three orthogonal pairs of parallel lines [7]. The calibrated camera allow complete rectification of the corresponding plane and determination of its orientation relative to the camera. By manually supplied parallel lines, Liebowitz et. al. [41] compute the vanishing line and a vanishing point for the orthogonal direction of a selected reference plane. Absolute 3D positioning of points and feature grouping are achieved after metric rectification of the reference plane. In [73], similar approach is proposed with the difference that reconstruction is formulated as a minimization process over the sum of squared distances for pairs of planes and points or lines.

Recently, an approach for reconstructing free-form, texture-mapped 3D scene models from a single painting or photograph is presented in [83]. The authors formulated the 3D generation as a constrained variational optimization problem. Given a sparse set of user-specified constraints on the local shape of the scene, a smooth 3D surface that satisfies the constraints is generated. Another similar method in [56]
used a closed-form method incorporating user-specified data such as surface normals, interpolation and approximation constraints to reconstruct a smooth surface from its image apparent contour, including multilocal singularities. Alternatively, in [60] a supervised learning approach is proposed, in which a set of monocular images of unstructured indoor and outdoor environments are adopted for training. Supervised learning is then applied to predict the value of the depth map as a function of the image by a hierarchical, multiscale Markov Random Field that incorporates multiscale local- and global-image features, and models the depths and the relation between depths at different points in the image.

Reconstruction of 3D models from a single image often requires more information about the scene or the object to be reconstructed, and generally produce only limited simplified representations of the scene as it is seen from the specific viewpoint. The occlusion ambiguity problem encountered in these approaches is generally addressed separately, which undermines the recovery performance for true 3D shape generation.

1.2 Multi-view approaches

Multi-view approaches provide solutions to resolve occlusion ambiguities and draw more attention from researchers. The general idea covering these methods is that given several images which sufficiently cover the whole object or the scene, a representation of its 3D shape can be computed. Among all multi-view approaches, volumetric methods gain more popularity due to the relative simpleness and robustness of the computational models. These methods project each point from a 3D grid to each image using known camera matrices and measure consistency of the projections. In the case a 3D point consistently projects to matching locations across images, it is
marked as a “scene point” (or opaque), otherwise it is marked as “transparent”. The information of color or grey scale variance can be considered as constraints in the 3D reconstruction, under the assumption of constant illumination and Lambertian reflectance. This constraint is known as photo consistency and is the basis of a large body of techniques on volumetric reconstruction. Silhouette-based approaches, on the other hand, abandon color information by segmenting object from its background and create binary images, hence the limitations of photo consistency measures are eliminated. The complete 3D model is recovered by intersection of all silhouette cones back-projected from the camera centers. The reduced computational complexity of these methods make them a common choice in practical applications. In the following sections, a few well-known approaches, including structure from motion, plane sweep stereo, voxel coloring and space carving, as well as some silhouette based algorithms are briefly discussed.

1.2.1 Structure from motion

In the field of computer vision, automatic feature extraction and matching are two well studied problems. Assuming the position of a point in 3D is somewhere along the ray connecting the camera optical center and the corresponding spot in the image plane, tracking its projections across multiple images and using triangulation allows the relatively accurate localization of the point in 3D. In such cases, the correspondences between the images and the reconstruction of 3D object need to be found. If extraction and correspondence can be performed for a sufficient number of features such as corner points and lines over images acquired from different locations, then both the 3D locations of the features and the camera positions can be simultaneously
estimated. This reconstruction method is called structure from motion (SFM). An extensive literature review of this methodology can be found in [52].

A method proposed by Morris et al. [51] aimed at scene topology estimation assuming only a sparse set of 3D point estimates computed from a SFM process. An initial triangulation of this point set is computed, then the resulted model is projected onto each image in the sequence. The problem of finding a consistent triangulation is formulated as a minimum variance estimation problem over the space of all possible triangulations. Despite the good surface approximations the approach may produce, for complex real scenes it carries an intrinsic disadvantage that no occlusions can be handled.

The focus placed on the approximation of the scene geometry fails to address the problem of reconstructing the surfaces that connect the available 3D features. Solutions provided by [5], [20] resulted in simplified and unconnected models while attempts to build a continuous triangulated model by a single view planar triangulation method [23] are only valid in the absence of occlusions. Recent development of this method can be found in [14], in which all possible assignments of 3D-points to image features are considered and the correct correspondence is established through an iterative expectation-maximization scheme where the E-step computes assignment weights and the M-step structure and motion parameters. Similar approach which also claims that no correspondence is required is proposed in [46]. An application of SFM for automatically reconstruct architecture is extended by [16]. The method explicitly makes use of strong geometric constraints such as perpendicularity and verticality which are likely to be found in architecture. The 3D scene structure is presented as a piecewise planar model which is initialized automatically by segmenting
a feature-based reconstruction. However, for more complex structures, the method requires more specific priors on building shape which may not be available. In [84] the authors proposed an algorithm to achieve dense surface reconstruction by iteratively estimating affine camera parameters, illumination, shape, and albedo in an alternating fashion.

1.2.2 Plane sweep stereo

Another line of research on scene recovery exploits 3D planes to reduce complexity of visibility tests in volumetric methods. These methods, which are referred to as “plane sweep stereo” which was first introduced in [9], consider a series of planes related by 3D homographies in the object space. When pixels from different views are back-projected, their intersections with these planes are analyzed for consistency to estimate their 3D positions. The consistency of back-projections and their intersections at a particular grid on a plane can be computed by a voting or photo-consistency mechanism. Voting based approaches back-project edges [9] and silhouettes [77]; while photo-consistency based methods, such as in [22] and [48], use the absolute difference of intensities as their dissimilarity measure. Plane-sweep stereo is also used to compute dominant scene planes, such as building facades, by measuring their consistency through homographies between the images induced by the sweep-planes [79]. This approach found its capable application in architecture reconstruction for its ability to model planar facets. In [3], an automatic plane reconstruction method utilizing plane sweep strategy is presented. The authors proposed a batch approach where 3D line positions are estimated by tracking their projections over successive frames. The
plane sweep is then applied by hypothesizing a planar facet attached to each line for every angle around it.

### 1.2.3 Voxel coloring and space carving

Voxel coloring methods generalize plane sweep approaches by replacing plane sweep with surface sweep [62]. A volume enclosing the whole real world scene is defined and is subdivided into voxels. Every voxel centroid is projected onto each of the images and photo consistency check is performed to determine if it is transparent or opaque. A threshold is set for the variance of the colors of the associated pixels in these images. However, the success of photo consistency check rely on precise camera calibration and is sensitive to noise. In addition, this test has to be performed for every voxel on all the images from where it can be visible, which can be a very computationally expensive process.

In order to address the problems of photo consistency, in [34] the authors have presented the “space carving” approach which generates an initial reconstruction that envelops the object to be reconstructed. The surface of the reconstruction is then eroded at the points that are inconsistent with the input images. The approach yields a unique reconstruction called the photo hull. The photo hull does not provide a minimal reconstruction but one which ensures that no valid voxels are disregarded from the consistency test. In [11], another alternative method called generalized voxel coloring (GVC) has been proposed. GVC utilizes layered depth image (LDI) data structure which associates every pixel with a linked list of voxels sorted in depth order. Invalid voxels are carved out and replaced with subsequent voxels in the corresponding
LDI entry by iterative photo consistency test. GVC is applicable to cases where cameras are placed outside the voxel space and provides faster convergence.

These methods collectively assume the object surface has lambertian properties and is illuminated by constant illumination. Due to the imaging geometry, multiple 3D points may project to the same image pixel and satisfy the photo consistency measure. This condition results in inaccuracies in the recovery process. In order to avoid such inaccuracies, these methods perform a visibility test for each scene point by, for instance, comparing surface normals in its vicinity. Visibility test can also be improved by statistical reasoning for grids which are occluded in some of the views [31]. In [35], the author has proposed the relaxation of the photo consistency definition, which results in robust reconstruction under calibration and image noise. For the relaxing of constant illumination assumption adopted by all the voxel coloring methods, Saito et al. [59] presented a method that requires only the fundamental matrices that relate every image.

1.2.4 Shape from silhouettes

A 2D silhouette is a representation of an object projected onto the image plane. The interior region corresponding to the projected object outline is featureless and is usually expressed by either black or white color, or is labeled with one or zero (binary image). Typical shape from silhouettes techniques start with acquiring images of the object from different views. The object silhouette can be extracted manually, interactively or automatically using simple differencing or segmentation. A volume bounding the object is then back-projected to 3D space from each silhouette. The object shape is then defined as the intersection of these volumes associated with
the set of acquired images. In the context of silhouette based methods, the photo consistency and visibility tests are eliminated by generating a visual-hull [39]. When the number of images used increases, visual hull generates a better fit to the actual object volume. However, in [40] it is proved that this number can be unbound for reconstruction of general polyhedral objects.

Silhouette based approaches reveal their inability in reconstructing non-convex objects, since concavities in the object geometry result in self-occlusions and cannot be resolved unless additional information about the occlusions is provided. Another factor that can significantly affect the performance of these approaches is the positioning of the cameras. Ambiguities arise where a complex object is imaged from a small number of locations. Taking into account some prior knowledge on the object shape, an iterative method presented in [65] specifies the viewpoints that optimize the reconstruction.

If the purpose is to approximate a crude model of an object, shape from silhouettes act as a particularly good approach for fast 3D generation. The methodology is intuitive and easy to implement. Recent developments and applications for shape from silhouette can be found in [50] [68] [26]. Specifically in [8], it was suggested that shape from silhouette across time and camera geometry provide sufficient information for reconstructing the object shape without establishing point correspondences. The shape can even be recovered from incomplete silhouettes provided that object motion between two frames is correctly estimated [75]. The most recent results in [43] and [21] show that this track remains active and there is still space for improvement.
1.3 Proposed approach

The main limitation of the aforementioned approaches is the requirement for precise camera poses and calibrations. While having pose and calibration is a desired property, it is, however, not an intuitive one. Explicit calibration is not always practical and has to be repeatable. In addition, methods for estimating pose and calibration parameters require dense and sub-pixel point correspondences as well as their initial estimates. Both of these requirements may not be practical due to projective changes across the views and appearance variation due to different lighting conditions. An alternative possibility is to perform auto-calibration in the recovery procedure, which utilizes the rigidity of the scene to obtain constraints on the camera parameters. In [80], perspective reconstruction of deformable structures is obtained from an uncalibrated image sequences. Assuming multiple objects are moving with constant velocity in the scene, Han and Kanade [27] proposed an approach to recover camera motion and camera parameters simultaneously and then reconstruct the scene by forcing metric constraints. In [55], the metric 3D reconstruction of both structure and motion is achieved for the image sequences containing common features located on a plane. While these approaches provide acceptable scene recovery using uncalibrated image sequences; they, however, require estimation of camera intrinsic parameters as an intermediate process. This requirement limits their applicability to cases where these parameters cannot be recovered. These limitations motivated development of shape recovery methods from images with no calibration and pose information.

In this dissertation, a computationally efficient and intuitive shape recovery approach is developed for cases where cameras have no calibration or pose information. The idea inspiring the proposed research is adopted from tomography imaging which
provides a stack of pseudo cross-sectional images of inspected organs [58]. Following the same line of thought, a set of hypothetical planes in the object space, which create cross-sections of objects, is considered. These planes generate 2D discrete subspaces with respective 1-1 mappings between images. Using these hypothetical planes, Khan et al. [33] demonstrated that the 3D scene structure can be recovered from “uncalibrated affine-cameras” when a set of three orthogonal vanishing points are known. We have recently demonstrated a method for “uncalibrated projective-camera” case when multiple [37] [36] or one vanishing point in the plane-normal direction is available [38].

In the following chapters we

1. provide a more detailed discussion of the ideas,

2. exploit different subspace topologies,

3. provide quantitative analysis of the recovered 3D in case of noisy observations, and

4. demonstrate the effectiveness of discussed methods with both qualitative and quantitative results by comparing recovered 3D against the ground truth.

One may find analogies of the proposed approach to plane sweep stereo and other volumetric methods. Specifically, plane sweep stereo methods also utilize plane projective geometry to explicitly model the 3D. In contrast to plane sweep stereo, we consider the inverse problem and provide analytical relations between the images with respect to a set of hypothetical planes. A significant difference to other methods is that these relations implicitly model the geometry of planes and their images and eliminate the requirement for camera calibration and pose.
1.4 Overview

In this chapter we have discussed some related works and their limitations which motivate the proposed approach. The remainder of the dissertation is organized as follows:

In chapter 2, the fundamental concepts in projective geometry as they relate to the proposed research, including vanishing point, vanishing line, and homography transformation, are briefly reviewed. The 3D information implied by vanishing point and vanishing line, when exploited, constructs the basis for the proposed approach. The homography transformation induced by a plane provides simple yet geometrically strong constraint which relates the object space and the image space.

The concepts and principles introduced in chapter 2, when extended, enable estimation of points lying on a set of hypothetical planes. In chapter 3, topology of these planes for two different subspaces in the 3D projective space is introduced. For each plane in either subspace, the image of a set of points on the plane is estimated by utilizing the scale factors associated with the points. Mathematical derivations, which give rise to the equations that establish point correspondences with respect to these hypothetical planes are provided in this chapter.

From the corresponding points estimated in chapter 3, homographies relating multiple images of a scene can be computed. These planar transformations are incorporated in a simple procedure for warping images to create object outlines. Based on the choice of the topology, details on how 3D is recovered are elaborated in chapter 4. Concerns on the limitations and dominant error of the approach are also addressed in this chapter.
Chapter 5 is dedicated to several experiments, which are conducted using data sets with different characteristics to prove the effectiveness of proposed approach. Qualitative and quantitative comparisons reveal that the proposed approach is capable of generating shape approximation efficiently. Finally, Chapter 6 concludes the research and points out the possible directions for future work.
CHAPTER 2

Fundamentals of plane projective geometry

The process of image formation can be generally described from two different aspects: the geometry which determines where in the image plane the projection of a point in the scene will be located, and the physics of light which determines the brightness of a point in the image plane as a function of illumination and surface properties. In this research, the focus is placed on exploiting the geometrical relations between the images of interest objects in the scene. More specifically, we conjecture that object shape can be recovered by exploiting the plane projective geometry between different views of a scene induced by a stack of hypothetical 3D planes. The geometry of these planes and the images provides 1-1 mappings between the points across images. In order to better manifest the core techniques adopted by the proposed method, the following discussion briefly introduces some important concepts and projective geometric relations between the images with respect to a single 3D plane.

2.1 Homogeneous Coordinates

In the Euclidean 2-space, a point \( x \) in \( \mathbb{R}^2 \) may be represented in its vector form as a pair of coordinates \((x, y)^T\). Adding an extra coordinate to this pair, which
becomes a triple \((x, y, 1)^T\) in \(\mathbb{P}^2\) space, results in the homogeneous coordinates of the point. Furthermore, \((kx, ky, k)\) denotes an equivalence class of coordinate triples for any non-zero value \(k\) to represent the same point. One of the major purposes of using homogeneous coordinates is to capture the concept of infinity. In the Euclidean coordinate system, infinity does not have a representation as a finite entity, while in homogeneous coordinates a point at infinity, or the so-called ideal point, is simply represented by \((x, y, 0)^T\).

A line lying on the x-y plane is represented by the equation \(ax + by + c = 0\), where different choices of \(a\), \(b\), and \(c\) give rise to different lines. If the line \(l\) is represented in the vector form as \(l = (a, b, c)\), the point \(x = (x, y, 1)^T\) lying on the line satisfies \(l^T x = 0\). Similar to the homogeneous point representation, since the line equations \(l^T x = 0\) and \((kl)^T x = 0\) are the same for any non-zero \(k\), \(l\) and \(kl\) are considered the same line in the projective space. Any homogeneous vector \((a, b, c)^T\) is a particular representative of the equivalent class. For a point \(x = (x, y, 0)\) lying at infinity, the line passing through \(x\) is in the form of \(l = (0, 0, 1)\). In fact, every ideal point lies on the specific line, which is called the line at infinity.

Given two lines \(l\) and \(l'\), from the triple scalar product identity we have \(l \cdot (l \times l') = l' \cdot (l \times l') = 0\) (\(\cdot\) is the vector inner product and \(\times\) is the cross product). If the vector \(x = l \times l'\) is thought to be representing a point, it is observed that \(l^T x = l'^T x = 0\). Hence, \(x\) must lie on both lines and is actually the intersection of the lines \(l\) and \(l'\). By an analogous argument, the line \(l\) passing through two points \(x\) and \(x'\) is obtained from \(l = x \times x'\). One may easily verify that \(l^T x = l^T x' = 0\). Another observation reveals that, since \(l^T x = 0\) implies that \(x^T l = 0\), the positions of line and point are swappable. This property leads to a general principle as described below:
Duality principle. For any theorem of 2-dimensional projective geometry there exists a dual theorem, which can be derived by interchanging the role of points, lines and planes in the original theorem. For instance, in the projective plane, any two distinct points define a line. Dually, any two distinct lines define a point (their intersection). Note that duality only holds universally in projective spaces. For other spaces such as in the affine plane, parallel lines remain parallel and do not intersect at all [28]. Once the transformation between points is established, the duality principle is applied to estimates the transformation between lines.

Homogeneous representation of points and lines constitute a basis for the projective geometry used extensively to unify the operations of common projection from a three-dimensional scene onto a two-dimensional image plane, or from one image onto another. Throughout this dissertation, the homogeneous coordinates are used for points both in the image and object spaces. In similar fashion to the 2D Euclidean space, the homogeneous coordinates for a point in 3D projective space $\mathbb{P}^3$ is represented as $(X, Y, Z, 1)^T$. The extension can be applied to the coordinates in any dimension, but in this report we only discuss 2D and 3D representations.

2.2 Projection from 3D object space to 2D image

The location of a point (or pixel) in the image space, which corresponds to a point in the object space, is determined from the geometry relating 3D to 2D. In order to describe this relation, the pinhole camera geometry shown in Figure 2.1 is widely adopted as the general projective camera model. Let a point $X$ in the object space be projected onto a pixel $x$ on the image, then the projection model is represented as
a linear transform:

\[ x = PX \]  

where \( P \) is called the projection matrix. It is realized that \( P \) is composed of internal and external parameters of the camera. The expression of \( P \), when represented with respect to these parameters, is

\[ P = KR[|I|\hat{C}] \]  

where \( K \) is the calibration matrix corresponding to the transformation from image coordinate frame to pixel coordinate frame. The elements included in \( K \) may vary by the intrinsic characteristics of the camera but is generally in the form of

\[ K = \begin{bmatrix} f_x & \alpha & x_0 \\ f_y & y_0 & 1 \end{bmatrix} \]
where \( f_x \) and \( f_y \) represent the focal length of the camera in terms of pixel dimensions in the x and y directions respectively. The skew parameter \( \alpha \) is usually set as zero for digital cameras. The parameters contained in \( K \) corresponds to the internal camera parameters, while external parameters describing the attitude and position of the camera are represented respectively by the rotation matrix \( R \) and translation vector \( \tilde{C} \). The \( 3 \times 3 \) matrix \( R \) contains the rotations with respect to the angles between world and camera coordinate frames, and vector \( \tilde{C} \) represents the coordinates of the camera center in the world coordinate frame. From the description above one may notice that, prior to determining \( P \) explicitly, these internal and external parameters need to be retrieved. Though rigorous procedure for camera calibration and orientation can be established in a well-controlled environment such as a laboratory, it is somehow impractical and time-consuming when attempting to estimate \( P \) for every image in a large collection. In the proposed approach, the 3D-2D projection is reduced to 2D-2D transformation which can be estimated from as few as only four correspondences, hence the practicability and efficiency is improved.

### 2.3 Vanishing Point and Vanishing Line

A distinguished feature in the projective geometry is that the object stretched out to infinity can have a finite extent in the image space. This feature in fact provides implied information about the 3D space which is exploited in the proposed approach to enable the object shape recovery. In order to realize the complete framework of the approach, a brief description of the concepts of vanishing points and vanishing lines will be provided in the following sections. For a more detailed discussion, readers are referred to \[28\].
2.3.1 Vanishing Point

A vanishing point in the image space is referred to as the image of an ideal point lying at infinity. Given a line pair parallel in the object space, the vanishing point can be measured by intersecting these lines in the image coordinates, as demonstrated in Figure 2.2. More formally, let two pixel pairs \((x_1, x_2)\) and \((x_3, x_4)\) lie respectively on images of two lines \(l_1\) and \(l_2\) parallel in the object space such that \(l_1 = x_1 \times x_2\) and \(l_2 = x_3 \times x_4\). The vanishing point \(v\) corresponding to the ideal point lying at intersection of these lines in the object space is computed from their cross product by \(v = l_1 \times l_2\). In addition to this intuitive method, a large body of literature has started on automatic detection of vanishing points after Barnard [4] introduced the

Figure 2.2: Vanishing points and vanishing line.
use of the Gaussian Sphere as an accumulation space. Other approaches can be found in [61] [6] [15].

A vanishing point depends only on the direction of the lines, which means despite their positions all parallel lines with the same directions intersect at one single point. Though parallelism is not preserved after projection, the information about orientation implied by the vanishing point still plays the role of a key to retrieve camera parameters and 3D object shapes. For instance, in [25] vanishing points are utilized for automatic camera calibration. As a matter of fact, the relation can be easily observed from analyzing the projection matrix P. Let the projection matrix P be represented as \( P = [p_1\ p_2\ p_3\ p_4] \) where \( p_i \) denotes the \( i^{th} \) column vector of P. By definition, the ideal point of the X direction is expressed as \([1, \ 0, \ 0, \ 0]\). The image of this ideal point is obtained by:

\[
v_x = P \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = p_1, \tag{2.4}
\]

which is the first column vector in P. Accordingly, vanishing points of Y and Z directions correspond to the second and third columns of P. In other words, vanishing points of the directions corresponding to the three orthogonal axes are enclosed in the projection matrix, with possibly different scales. The inference that the fourth column of P represents the image of the origin in world coordinate frame can be obtained in a similar fashion.

2.3.2 Vanishing Line

Analogous to the idea that parallel lines in projective 2-space intersect at a vanishing point, parallel planes in projective 3-space, when extended to infinity, intersect
at a common line which is called the vanishing line of the plane. The most comprehensible example of such line is the horizon which is frequently observed in images. A vanishing line \( l_v \) is the image of the line at infinity passing through ideal points; hence, it can be estimated by the cross product of two vanishing points as \( l_v = v_1 \times v_2 \).

Vanishing line can also be estimated from equally spaced coplanar parallel lines, orthogonality relationship among vanishing points and lines, and by using the elation transform estimated from image texture [66]. Alternatively, vanishing line can be obtained from the geometry of the camera center, the vertical vanishing point and the principal point, which can be assumed to lie at the center of the image. This geometry is adopted by the proposed approach as the basis of 3D generation. The illustration and discussion will be provided in chapter 3, and more detail of related geometry can be found in [28]. Automatic detection for vanishing points and lines have also been proposed in [61] and [1] by grouping or clustering.

2.4 Homography Transformation

Let an image pixel, \( x \), and a 3D point, \( X \), be represented in the homogeneous coordinates using their canonical forms, such that \( x = [x, y, 1]^T \) and \( X = [X, Y, Z, 1]^T \).

In this setting, the projection from point \( X \) to pixel \( x \) is expressed by a linear mapping:

\[
\lambda x = PX = \begin{bmatrix}
p_1 & p_2 & p_3 & p_4
\end{bmatrix}
\begin{bmatrix}
X \\
Y \\
Z \\
1
\end{bmatrix},
\]

where \( \lambda \) is the scale factor due to projective equivalence of \( x \sim \lambda x \), \( P \) is \( 3 \times 4 \) camera projection matrix and \( p_i \) is the \( i^{th} \) column vector of \( P \). With the freedom to choose the world coordinate frame, we may select any scene plane \( \pi \) as the X-Y plane, such that the points on this plane have \( Z = 0 \). The image of a point on \( \pi \) is then given by
the plane projective transformation:

\[
 sx_{\pi} = H_{\pi}X_{\pi} = \begin{bmatrix} p_1 & p_2 & p_4 \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix},
\]

(2.6)

where \( s \) is a scale factor different from \( \lambda \) in (2.5) due to \( Z \) being 0. In this equation, the homography matrix \( H_{\pi} \) serves as a linear mapping between the points lying on the X-Y plane and the corresponding image pixels.

Considering two cameras that image the 3D scene plane, \( \pi \), at two different viewpoints, the homography transformation between the views has the following form:

\[
 sx_{\pi_i} = H_{\pi_i}X_{\pi} = H_{\pi_i}(H_{\pi_j}^{-1}x_{\pi_j}) = H_{ij}x_{\pi_j},
\]

(2.7)

where \( H_{ij} \) is the homography matrix between views \( i \) and \( j \) (see Figure 2.4 for illustration).

Figure 2.3: Homography between two images induced by a scene plane \( \pi \).
2.4.1 Estimation of planar homography matrix

Given a set of 2D to 2D correspondences between two images \( \{ x_{nj} : n = 1, 2, ..., N \} \leftrightarrow \{ x_{ni} : n = 1, 2, ..., N \} \) where \( N \geq 4 \), the homography transformation between them is expressed by the equation

\[
x_{ni} = Hx_{nj}.
\] (2.8)

Alternatively, the equation may be expressed by the vector cross product as

\[
x_i \times Hx_j = 0.
\] (2.9)

Let the transpose of \( m-th \) row vector of \( H \) be denoted by \( h^m \), then the cross product expression gives a set of three equations in the form of

\[
\begin{bmatrix}
0^T & -x_{nj}^T & y_{nj}^T \\
x_{nj}^T & 0^T & -x_{ni}x_{nj}^T \\
-y_{ni}x_{nj}^T & x_{ni}x_{nj}^T & 0^T
\end{bmatrix}
\begin{bmatrix}
h^1 \\
h^2 \\
h^3
\end{bmatrix} = A_n h = 0
\] (2.10)

However, only two of them are linearly independent since the third row can be represented as the linear combination of the other two rows. For the \( N \) point correspondences, each one provides two equations and when stacked, gives the equation:

\[
\begin{bmatrix}
0^T & -x_{nj}^T & y_{nj}^T \\
x_{nj}^T & 0^T & -x_{ni}x_{nj}^T \\
\cdots & \cdots & \cdots \\
0^T & -x_{Nj}^T & y_{Nj}x_{Nj}^T \\
x_{Nj}^T & 0^T & -x_{Ni}x_{Nj}^T
\end{bmatrix}
\begin{bmatrix}
h^1 \\
h^2 \\
h^3
\end{bmatrix} = Ah = 0
\] (2.11)

In the equation \( A \) is a \( 2N \times 9 \) matrix. The homography matrix has eight degree of freedom associated with rotation, translation, skew and a vanishing line, hence \( A \) has rank eight and has a 1-dimensional null space which provides a solution for \( h \).

The procedure described above, which is known as the basic Direct Linear Transformation (DLT) algorithm, is summarized as follows.
The basic DLT algorithm for 2D homography

1. For each correspondence $x_{ni} \leftrightarrow x_{nj}$, construct $A_n$.

2. Assemble the $2N \times 9$ matrix $A$ for all the $N$ correspondences.

3. Compute the Singular Value Decomposition (SVD) for matrix $A$ such that $A = UDV^T$ in which $D$ is diagonal with positive and descending diagonal entries. The singular vector corresponding to the smallest singular value of $A$ is the solution $h$, which is obtained from the last column of $V$.

4. The homography matrix is determined from rearranging $h$.

In the practical DLT computation, data normalization is considered an essential step to ensure better result. The normalization involves a transformation matrix consisting of a translation and scaling. This matrix transforms original point set to a new one such that the centroid of the new points is the coordinate origin $(0, 0)^T$, and their average distance from the origin is $\sqrt{2}$. In the presence of noise, data normalization avoids the effect of a large condition number in $A$, which may amplify the divergence between estimated and correct result. The following procedures constitute normalized DLT algorithm.

The normalized DLT algorithm for 2D homography

1. Given point correspondences $\{x_i \leftrightarrow x_j\}$, compute the similarity transformations $T_i$ and $T_j$ such that $\tilde{x}_i = T_i x_i$ and $\tilde{x}_j = T_j x_j$ are the normalized point sets.

2. Apply basic DLT to the correspondences $\{\tilde{x}_i \leftrightarrow \tilde{x}_j\}$ to obtain a homography $\tilde{H}$.

3. Set $H = T_i^{-1} \tilde{H} T_j$.

When considering the image region in which the homography is computed is small or the image has been acquired with a large focal length, an affine homography is
a more appropriate model of image displacements. An affine homography is a special type of a general homography whose last row is fixed to \( h^3 = [0, 0, 1] \). Based on this, the equations listed above can be further simplified. The homography can be alternatively estimated from three point correspondences and an epipole [28]. In some literature, homography can even be estimated automatically for particular applications from known scene structure. For instance in [67], homography relating the points and lines lying on the road is computed after estimating a dominant vanishing point.

### 2.4.2 Some properties of homography

After estimating homography which maps two point set as \( x_i = Hx_j \), the transform between corresponding lines \( l_i \) and \( l_j \) respectively lying on same planes as \( x_i \) and \( x_j \) is also determined. Assume points \( x_{1j}, x_{2j} \) and \( x_{3j} \) lie on \( l_j \), thus \( l_j^T x_{nj} = 0 \) for \( n = 1, 2, 3 \). When \( x_{nj} \) is mapped onto \( x_{ni} \) by \( H \), the line \( l_i \) passing these three points also fulfills the equation \( l_i^T x_{ni} = 0 \) for \( n = 1, 2, 3 \). By verifying that \( l_i^T H^{-1} H \), one comes to the inference that \( l_i = H^{-1} l_j \).

Another property regarding the configuration of point correspondence is of concern to the homography estimation. In order to compute homography between two point sets, a minimal requirement is the existence of four correspondences. However, even if this requirement is fulfilled, the success of homography estimation is not guaranteed. Suppose that three of the four points are collinear, the homography is not sufficiently constrained. As a result, there will exist a family of homographies mapping one point set to another. The situation where a specific solution of \( H \) is not determined from the configuration of correspondences, is termed degenerate condition. The condition
can be addressed prior to the estimation by testing if a point lie on the line linking any other two points.

2.5 Summary

In this chapter the definitions and estimations of vanishing point, vanishing line and homography transformation induced by a plane are introduced. The discussion covers merely a small fraction of the full scope of projective geometry. Nevertheless, it provides sufficient background knowledge for the proposed approach which we will elaborate in the following chapters. It is worth noting that whenever a scale factor associated with a point is presented, the last element in the homogeneous representation of the point is set to be one. Otherwise, the point is assumed to be expressed arbitrarily in its equivalent class.
CHAPTER 3

Topology of Subspaces in $\mathbb{P}^3$

The plane projective geometry provided in Chapter 2 constitutes the relations between images of corresponding points on a scene plane. Now consider that the 3D scene consists of infinite number of planes with arbitrary orientations. A set of such planes, when grouped by a specific arrangement or parameterization, form a subspace in the projective space. The topology of these planar discrete subspaces plays an important role in recovering the 3D object shape. In this chapter, we first introduce the general concept of parameterized planes, followed by the discussion on two possible parameterizations, their effectiveness and limitations in the recovery process.

3.1 Parameterizations of planes

Let there be a set of finite points $\{\mathbf{X}_n : n = 1, 2, ..., N\}$ lying on a reference plane $\pi_0$ in $\mathbb{P}^3$. When this plane is chosen to coincide with the X-Y plane in a local coordinate frame, it is represented as $\pi_0 = [0, 0, 1, 0]^\top$, and the homogeneous coordinates of any point in $\mathbf{X}_n$ is in the form of $[X_n, Y_n, 0, 1]$. Assume another point set $\{\mathbf{X}'_n : n = 1, 2, ..., N\}$ lying on a plane $\pi$ distinct from $\pi_0$ satisfies the mapping $\mathbf{X}'_n = H\mathbf{X}_n$ where $H$ is considered the $4 \times 4$ homography matrix. By definition,
\[ X_n^\top \pi_0 = 0 \] and \[ X_n^\top \pi = 0 \], and the transformation from \( \pi_0 \) to \( \pi \) related by \( H \) is simply \( \pi = H^{-\top} \pi_0 \). In another interpretation, plane \( \pi_0 \) serves as a generator for any arbitrary plane \( \pi \) in \( \mathbb{P}^3 \), such that an expression of \( \pi \) can be obtained from \( \pi_0 \) even if no physical plane corresponds to \( \pi \) in the scene. The homography matrix \( H \) which satisfies the above relations can be specifically manipulated to create unique subspace in \( \mathbb{P}^3 \). In the following discussion, we analyze two possible subspace topologies, where \( H \) either translates or rotates the points on the reference plane \( \pi_0 \) to a hypothetical plane \( \pi \).

### 3.1.1 Parallel Plane Topology

The ground plane, or X-Y plane \( \pi_0 = [0 0 1 0]^\top \) in the object space, when translated along the normal direction of \( \pi_0 \) by a distance \( Z \), as shown in Figure 3.1.1, creates a hypothetical plane \( \pi' = [0, 0, 1, -Z]^\top \) which is parallel to \( \pi_0 \). The vertical projection of a point \( X = [X, Y, 0, 1] \) from \( \pi_0 \) onto \( \pi' \) is the point \( X' = [X, Y, Z, 1] \), and it can be easily verified that \( \pi^\top X = 0 \) and \( \pi'^\top X' = 0 \). This translation process results in planes \( \pi' (Z) \) parameterized by \( Z \), such that the homography transformation between

Figure 3.1: Subspace topology for the parallel planes.
the points on $\pi_0$ to the points on $\pi'(Z)$ becomes:

$$X' = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & Z \\ 0 & 0 & 0 & 1 \end{bmatrix} X = H_{\text{trans}} X.$$ (3.1)

where $H_{\text{trans}}$ denotes the homography of translation. The collection of $\pi'$ for a set of specified $Z$ values forms a subspace in $\mathbb{P}^3$, and any point in $\mathbb{P}^3$ must reside in one of the planes in such subspaces. Under this configuration, the general $\mathbb{P}^3 \mapsto \mathbb{P}^2$ projection is then reduced to the $\mathbb{P}^2 \leftrightarrow \mathbb{P}^2$ planar homography as is described in Section 2.4.

### 3.1.2 Rotated Plane Topology

Following a similar strategy to the parameterization of parallel planes, hypothetical planes created by rotating the reference plane about an arbitrary axis provides a subspace topology. Let plane $\pi_0$ and point $X$ be defined as in the previous section.
A plane \( \pi_{\theta} \) in \( \mathbb{P}^3 \) which is not to parallel to \( \pi \) intersects \( \pi_0 \) at a line other than the line at infinity. From another point of view, \( \pi_{\theta} \) can be deemed the plane rotated from \( \pi_0 \) about the intersecting line by the included angle \( \theta \) between \( \pi_{\theta} \) and \( \pi \). For a point on \( \pi_0 \), its corresponding location on \( \pi_{\theta} \) after rotation is generally formulated with respect to three coordinate axes or an arbitrary 3D line with a common expression of \( X_{\theta} = RX \) where \( R \) represents the rotation matrix. However, the representation of rotation matrix can be very complicated and result in difficulties on formulating the relations. In order to simplify the formulation, a similar scheme as in the previous section is followed to find the projection of points on \( \pi_0 \) onto \( \pi_{\theta} \) along the Z direction.

Without lost of generality, one may select any line on the reference plane \( \pi_0 \) as the X axis. Assume \( \pi_{\theta} \) intersects \( \pi_0 \) at X axis and the included angle is \( \theta \). The vertical projection of \( X_{\pi} = [X, Y, 0, 1] \) onto \( \pi_{\theta} \) is denoted by \( X_{\theta} \) (see Figure 3.1.2 for illustration). From the simple geometry shown in the figure, the homogeneous coordinate of \( X_{\theta} \) is determined as \( X_{\theta} = [X, Y, Y \tan \theta, 1] \). Analogous to the parallel plane topology, there exist a homography \( H_{\text{rotate}} \) which maps \( X \) onto \( X_{\theta} \) as \( X_{\theta} = H_{\text{rotate}}X \) and is designated as:

\[
H_{\text{rotate}} = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & \tan \theta & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}.
\]  

(3.2)

The plane \( \pi_{\theta} = [0, \tan \theta, -1, 0] \) where the projected point \( X_{\theta} \) lie on, corresponds to one member in a pencil of planes. Hence, the subspace of rotated planes can be established by changing the value of \( \theta \) to generated a set of planes parameterized by \( \theta \). Note that the subspaces and accompanied homography transformations introduced in the section are merely two of the unlimited possible configurations specially designed for the use by proposed approach to recover object shape. By utilizing the particular
setting of either subspace and provided certain features identified on the images, the techniques introduced in the following sections are enabled to estimate point correspondences for any hypothetical plane.

**3.2 Points on hypothetical parallel planes**

The previous sections elaborate the constructions of two specific subspaces in $\mathbb{P}^3$. Now the follow-up task is how to locate the images of $X'$ and $X_\theta$ on any “hypothetical” plane. We should note that, since in general there may not exist physical features on these locations that are directly measurable on the images. Consider $\{X_n : n = 1, 2, ..., N\}$ is a set of points residing on the reference X-Y plane. In the image space their corresponding locations $\{x_n : n = 1, 2, ..., N\}$ are generated by projection equations given in (2.6). For the point set $\{X'_n : n = 1, 2, ..., N\}$ which is given by (3.1), its image can be written as:

$$\lambda_n x'_n = PH_{\text{trans}} \begin{bmatrix} X_n \\ Y_n \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} p_1 & p_2 & p_4 \end{bmatrix} \begin{bmatrix} X_n \\ Y_n \\ 1 \end{bmatrix} + Zp_3. \tag{3.3}$$

In this equation $p_3$ corresponds to the vanishing point for the $Z$ direction and can be replaced with $v_z$. Joining (3.3) with (2.6) results in a relation between conjugate pixels lying on the reference and hypothetical planes given by:

$$\lambda_n x'_n = s_n x_n + Zv_z. \tag{3.4}$$

Based on the fact that the last components of homogenous vectors are 1, the scale factor $\lambda_n$ in (3.4) is simply $\lambda_n = s_n + Z$. As suggested in equation (3.4), given the vanishing point $v_z$ and the scale factor $s_n$, any image point along the line passing through $x_n$ and $v_z$ can be estimated by setting $Z$ to different values based on the
parameterization. The vanishing point $v_z$ can be intuitively obtained from the image of a vertical parallel line pair, but the estimation of $s_n$ requires more delicate schemes under various situations, which are provided in the coming sections.

3.2.1 Scale estimation from parallelism in the object space

![Figure 3.3: Geometry of parallelogram and vanishing points.](image)

Refer to Figure 3.2.1, in which $\{X_n : n = 1, 2, 3, 4\}$ correspond to the corners of a parallelogram on the reference plane $\pi_0$ and its projection on the image is identified as $\{x_n : n = 1, 2, 3, 4\}$. Based on the configuration of parallel plane topology, there exists a point set $\{X'_n : n = 1, 2, 3, 4\}$ on the plane $\pi' = [0, 0, 0, -Z]$. The fact that parallelogram provides two parallel line pairs gives two vanishing points $v_1$ and $v_2$. In order to estimated the projection of $\{X'_n\}$, there has to be at least one point which is identifiable on the image. Let $x'_1$ be such point, then by rearranging equation (3.4)
as $s_1(x'_1 - x_1) = Z(v_z - x'_1)$, the scale factor for $x_1$ is computed by

$$s_1 = \frac{Z(x'_1 - x_1)^T(v_z - x'_1)}{(x'_1 - x_1)^T(x'_1 - x_1)}, \quad (3.5)$$

For other points $x'_2$, $x'_3$ and $x'_4$, the geometry shown in Figure 3.2.1 is applied for their retrievals. Take $x'_2$ for example, since $\overrightarrow{x'_1x_2}$ and $\overrightarrow{x'_1x'_2}$ are the images of parallel lines in the scene, and based on the fact that all parallel lines of same direction intersect at the same vanishing point, $\overrightarrow{x'_1x_2}$ intersect $\overrightarrow{x_1x_2}$ at $v_1$ in the figure. In addition, the $\overrightarrow{x_2x'_2}$ pass through $v_z$ because $\overrightarrow{X_2X'_2}$ is parallel to the Z axis. Hence, $x'_2$ lie at the intersection of $\overrightarrow{x'_1v_1}$ and $\overrightarrow{x_2v_z}$ and is can be computed by

$$x'_2 = (x_2 \times v_z) \times (x'_1 \times v_1) \quad (3.6)$$

Similarly, $x'_3$ and $x'_4$ are obtained as

$$x'_3 = (x_3 \times v_z) \times (x'_2 \times v_2)$$
$$x'_4 = (x_4 \times v_z) \times (x'_1 \times v_2). \quad (3.7)$$

Once the locations of points $\{x'_n : n = 2, 3, 4\}$ are estimated, their scale factors can also be determined from equation (3.5).

In the cases where $\{X_n\}$ gives only one parallel line pair, but instead there are two corresponding points of $\{X'_n\}$ identifiable in the image, such as the points $x'_1$ and $x'_3$ in Figure 3.2.1, a similar procedure is carried out to find $x'_2$ and $x'_4$ as

$$x'_2 = (x_2 \times v_z) \times (x'_1 \times v)$$
$$x'_4 = (x_4 \times v_z) \times (x'_3 \times v). \quad (3.8)$$

Note that in Figure 3.2.1, the reference plane does not coincide with the actual ground plane but set to the wall of the building instead. Though parallelism is commonly observed in the real world environment such as on the man-made buildings in the urban area, for a more generalized situation, the feature points on the reference plane
can be randomly distributed. In the next section a potential solution is introduced for tackling the problem of lacking parallelism.

### 3.2.2 Scale estimation from scale ratio

If the point set \( \{X_n\} \) is arbitrarily located on the reference plane \( \pi_0 \), it is possible that these points don’t belong to any parallel line pair, hence the approach discussed in the previous section can not be applied. However, under certain restrictions an approach which we refer to as the *scale ratio* in this dissertation is proposed to estimate the scale factor for each point.
Again let the height $Z$ for the endpoint $X'_1$ to be known or presumed, and $s_1$ accompanying $x_1$ is computed from (3.5). By exploiting the algebraic and geometrical relations between points, we propose the following lemma:

**Lemma 1. Scale ratio of two image points.** When projecting two points $X_1, X_2$ lying on a plane in the object space onto the corresponding points $x_1, x_2$ in the image space using homography, the ratio of the introduced scale factors $s_1$ and $s_2$ is the inverse proportion of distances from the image points to the vanishing point $v$ of $\overrightarrow{X_1X_2}$ direction. More specifically, let $D(x, v)$ denote the distance between two image points $x$ and $v$, then the scale ratio is expressed as:

$$\frac{s_1}{s_2} = \frac{D(x_2, v)}{D(x_1, v)}. \quad (3.9)$$

**Proof.** Given two points $X_1 = [X_1, Y_1, 1], X_2 = [X_2, Y_2, 1]$ on the ground plane $\pi_0$ and the corresponding image pixels $x_1, x_2$, equation (2.6) provides the following relation:

$$s_1x_1 - s_2x_2 = H(X_1 - X_2) = H \begin{bmatrix} X_1 - X_2 \\ Y_1 - Y_2 \\ 0 \end{bmatrix}. \quad (3.10)$$

The right hand side is equivalent to computing the vanishing point $v$ for the $\overrightarrow{X_1X_2}$ direction, which can be expressed as

$$H(X_1 - X_2) = kv, \quad (3.11)$$

where $k$ is a scale factor associated with $v$ under projective equivalence. As stated in the previous chapter, the last component of the homogeneous coordinates is set to be one. By comparing the coefficients of equations (3.10) and (3.11), we observe that $k = s_1 - s_2$. Thus, the two equations can be combined and rearranged as:

$$s_1(x_1 - v) = s_2(x_2 - v). \quad (3.12)$$
Taking norms to both sides of the equation above gives

\[
\frac{s_1}{s_2} = \frac{|x_2 - v|}{|x_1 - v|}
\]  

(3.13)

Each norm in equation (3.13) can be interpreted as the distance between two points, hence lead to the relation described in the lemma.

![Figure 3.5: Diagram for the scale ratio in Lemma 2](image)

The relation can be further extended by applying the property of similar triangles as shown in Figure 3.5 which leads to:

**Lemma 2.** The scale ratio in Lemma 1 equals to the inverse proportion of distances from the image points to the vanishing line \( l_v \) of these parallel planes:

\[
\frac{s_1}{s_2} = \frac{D(x_2, l_v)}{D(x_1, l_v)}.
\]

(3.14)

The proof of the lemma above is evident from Figure 3.5: The ratio of the distances from \( x_1 \) to \( v \) and from \( x_2 \) to \( v \), which is \( b/a \), equals to \( d/c \) according to similarity of
triangles. As long as the vanishing line is observed or computed from measuring two parallel line pairs, even if these lines don’t lie on the reference plane, the scale factor for any point in \( \{x_n\} \) can be estimated from scale ratio.

![Diagram](image)

Figure 3.6: The geometric relationship between the vanishing point of a reference direction and the vanishing line of the planes orthogonal to the direction.

Although the lemmas provide more freedom to the distribution of the feature points \( \{X_n\} \), the existence of lines parallel to \( \pi_0 \) is not always guaranteed, making the direct estimation for vanishing line infeasible. Nevertheless, we propose a novel way for the indirect estimation of vanishing line. By examining the geometry of a single image shown in Figure 3.6, one may notice that the vertical vanishing point \( v_z \) is the image of the projection of camera center on the reference plane \( \pi_0 \), and the vanishing line \( l_v \) lies on the pencil of image plane and the plane parallel to \( \pi_0 \) passing through the camera center. The line passing through the principle point \( pp \) and the vanishing point is obtained by \( l_{pp} = pp \times v_z \). Let this line intersect the vanishing line
at pixel $p$, such that $l_{pp}$ is perpendicular to $l_v$. This observation suggests that given $l_{pp} = [a, b, c]^{\top}$, the line perpendicular to it has the form: $l_v = [b, -a, d]^{\top}$, where $d$ is unknown so far. The orthogonal distance from pixel $x_n = [x_n, y_n, 1]^{\top}$ to the unknown vanishing line $l_v$ is denoted by $D(x_n, l_v)$ computed by:

$$D(x_n, l_v) = \frac{bx_i - ay_i + d}{\sqrt{b^2 + (-a)^2}}. \quad (3.15)$$

Assume the scales of two points in $\{x_n\}$, say $x_1$ and $x_2$, are known or estimated as $s_1$ and $s_2$. By integrating equation (3.15) into the scale ratio in lemma 2, we come up with the expression:

$$\frac{s_1}{s_2} = \frac{D(x_2, l_v)}{D(x_1, l_v)} = \frac{bx_2 - ay_2 + d}{bx_1 - ay_1 + d}, \quad (3.16)$$

from which the only unknown $d$ can be estimated as:

$$d = \frac{s_2(bx_2 - ay_2) - s_1(bx_1 - ay_1)}{s_1 - s_2}. \quad (3.17)$$

Note that for an uncalibrated camera, the real principle point remains undetermined. In computer vision literatures it is commonly assumed to sit on the image center. However, if the image $I'$ is a crop from an original frame image $I$, as shown in Figure 3.2.2, the vanishing line $l'_v$ estimated from scale ratio may be way off the actual one, $l_v'$, thus result in wrong scale factors.

The technique trade for the removal of two vanishing points with adding another predetermined scale factor for the purpose of retrieving vanishing line of $\pi_0$. The suitability of these alternatives should be judged upon the actual characteristic of available features in the scene.
3.2.3 Scale factors in multiple images

Discussions up to the point in this chapter are made for the geometry involving a single image. When multiple images of a scene are taken into consideration, additional techniques can be utilized to further expedite the computation and reduce vanishing line estimation error as well. Let $I_i$ and $I_j$ denote two images in which respective point sets $\{x_{ni}\}$ and $\{x_{nj}\}$ are the projections of a point set $\{X_n\}$ which lie on the reference plane $\pi_0$. The homography matrix mapping $\{x_{ni}\}$ onto $\{x_{nj}\}$ as $x_{nj} = H_{ji}x_{ni}$ can be estimated from the algorithm discussed in section 2.4. Once the vanishing line $l_{vi}$ in the image $I_i$ is determined, based on the duality principle it can be transformed to any other image $I_j$ by the homography as $l_{vj} = H_{ji}^{-T}l_{vi}$, in which $H_{ji}^{-T}$ is the transposed inverse of the homography from image $I_i$ to image $I_j$. 

Figure 3.7: False estimation of vanishing line for cropped image.
After the scale factors of all points in \( \{x_{ni}\} \) and one point in \( \{x_{nj}\} \) have been computed from scale ratio, for the corresponding points in \( I_j \), the scale factors can be obtained from the scale ratio across images as delineated in the following lemma:

**Lemma 3. : Scale ratio across images.** Known scale factors \( s_{1i} \) and \( s_{2i} \) for two image pixels \( x_{1i} \) and \( x_{2i} \) in image \( I_i \) provide the scale ratio for the conjugate pixels \( x_{1j} \) and \( x_{2j} \) in image \( I_j \) by:

\[
\frac{s_{1j}}{s_{2j}} = \frac{k_1 s_{1i}}{k_2 s_{2i}},
\]

where \( k_n = h_3 x_{ni} \) such that \( x_{nj} = \frac{1}{k_n} H_{ji} x_{ni} \) for \( n = 1, 2 \) and \( h_3 \) is the third row of the estimated homography \( H_{ji} \) which maps the images of points on reference plane \( \pi_0 \) from \( I_i \) to \( I_j \).

*Proof.* For the homography matrix \( H_{ji} \) induced by reference plane \( \pi_0 \), the duality theorem suggests the vanishing line \( l_{vj} \) transformed from \( l_{vi} \). Following the results in Lemma 2, the scale ratio across images can be shown to result in:

\[
\frac{s_{1j}}{s_{2j}} = \frac{l_{vj}^\top x_{2j}}{l_{vj}^\top x_{1j}} = \frac{\frac{1}{k_2} l_{vi}^\top H^{-1} H x_{2i}}{\frac{1}{k_1} l_{vi}^\top H^{-1} H x_{1i}} = \frac{k_1 D(x_{2i}, l_{vi})}{k_2 D(x_{1i}, l_{vi})} = \frac{k_1 s_{1i}}{k_2 s_{2i}}.
\]

Hence, the scale factors for the points on \( I_j \) are estimated.

One should note that due to the duality of homography, estimation of the vanishing line in only one image is adequate, consequently improves the computational time.

The use the relations introduced above determines the scale factors for image pixels corresponding to co-planar 3D points for different hypothetical planes residing at \( Z \). When multiple images of a scene are given, these pixels are used to estimate
the homographies induced by the hypothetical planes, which in turn provides cross-
sections of scene objects. Details on how to use the induced homographies for the
scene recovery is outlined in Chapter 4.

3.3 Images of points on rotated planes

Following the subspace configuration of a pencil of planes outlined in section 3.1.2,
the estimation for the image pixels of the points on each plane is carried out in a way
similar to that in the parallel plane topology. As defined in Section 3.1.2, a point
\( X_\theta \) on plane \( \pi_\theta \) relates to \( X \) on \( \pi_0 \) by \( X_\theta = H_{\text{rotate}}X \). The projection of \( X_\theta \) into the
image plane results in:

\[
\lambda_\theta x_\theta = PH_{\text{rotate}}X = \left[ \begin{array}{ccc} p_1 & p_2 & p_4 \end{array} \right] \left[ \begin{array}{c} X_n \\ Y_n \\ 1 \end{array} \right] + Y_n \tan \theta p_3.
\] (3.20)

Based on the fact that the first term on the right is the image of \( X \), the equation can
be further written as

\[
\lambda_\theta x_\theta = sx + Y \tan \theta p_3
\] (3.21)

in which \( s \) is the scale factor associate with \( x \) under projective equivalence. Rear-
arranging this equation and substituting \( v_z \) for \( p_3 \), the projection becomes:

\[
s_\theta x_\theta = x + ts_z v_z,
\] (3.22)

where \( s_\theta = \frac{\lambda_\theta}{s} \), \( t = \tan \theta \) and \( s_z = \frac{Y}{s} \). Considering the canonical form for the
homogeneous pixel coordinates, \( s_\theta \) in the above equation can be substituted with
\( s_\theta = (1 + ts_z) \). For different points on the plane, the scale factors \( s_z \) change with
their orthogonal distances to the \( X \) axis, which correspond to their \( Y \) coordinates.
In the following section, a geometric technique for solving $s_z$ without knowing the $Y$ coordinate for every point is presented.

### 3.3.1 Scale factor estimation

![Figure 3.8: Geometry of the pixels under rotated plane topology.](image)

Let a vertical feature identified in an image has two endpoints $x$ and $x_\theta$ respectively corresponding to scene points $X$ and $X_\theta$ (see Figure 3.1.2). In the case when the height $Z$ and distance to the axis of rotation $Y$ of a point known or presumed, such that $t = \tan\theta = Z/Y$ can be determined, the scale factor $s_\theta$ in (3.22) can be estimated from:

$$s_\theta = \frac{1}{t} (A^T A)^{-1} A^T B,$$  \hspace{1cm} (3.23)
where $A = v_z - x_\theta$ and $B = x_\theta - x$.

The scale computed from (3.23) for this point serves as a reference to estimate the scale factors $s_\theta$ for other points on the plane based on the geometry shown in Figure 3.8. Assume the vanishing point $v_z$ is estimated from vertical linear features, and $x_1$ and $x_2$ are projected from the points on plane $\pi$ and $x_{1\theta}$ is identified on the image as well. The line connecting $x_1$ and $x_2$ intersects line $l$, which is the image of the axis of rotation, at pixel location $m$. Since $m$ is on the axis, its position remain unchanged after rotation, which means the line $\overrightarrow{x_{1\theta}x_{2\theta}}$ still intersects the axis at $m$. Hence, $x_{2\theta}$ lie at the intersection of $\overrightarrow{x_2v_z}$ and $\overrightarrow{x_1m}$ and its image coordinates can be computed from the cross product:

$$x_{2\theta} = l_{x_1,m} \times l_{x_2,v_z},$$

(3.24)

where $l_{x_1,m}$ and $l_{x_2,v_z}$ are respectively the lines connecting $(x_{1\theta}, m)$ and $(x_{2\theta}, v_z)$. In this setting, (3.23) provides the scale factor for the second point, $x_2$.

Given a point set \{x_n : n = 1, 2, 3,...N\}, changing the value of $\theta$ in (3.21) provides new point sets for the corresponding plane $\pi_\theta$. Typically, one may freely select any two points on the reference plane to construct the axis of rotation, and the remaining $N - 2$ points are used along with $\theta$ to estimate their locations on the image related to rotated planes. However, the location of rotation axis may affect the recovery under certain conditions. These conditions and possible solutions are left for more detailed discussion in the next chapter.

### 3.4 Summary

The discussions provided in this chapter outline the concept of subspace topology behind the proposed shape recovery approach. Particular arrangements of planes
are devised and equations are derived accordingly to enable the estimation of points on hypothetical planes. Specific emphasis has been place upon the lemmas for the estimation of scale factors, which has not been revealed in any other literature. One may have noticed the simpleness of the computation involved in this chapter that is mainly composed of fundamental linear algebra. Indeed this characteristic realizes the construction of a fast recovery algorithm and is one of the merits of the approach.
CHAPTER 4

Shape Recovery

The topology of subspaces introduced in the previous chapter constitutes the core techniques of the proposed approach for shape recovery. In the following discussion we provide a detailed procedure for generating 3D object shape.

4.1 Shape representation

The proposed process for recovering a 3D scene starts with detecting coplanar pixel correspondences and vertical vanishing points. Let the images of a target object be \{I_m, m = 1, 2, 3, ...M\} where \(M\) is the number of images captured by multiple cameras at different viewpoints simultaneously. According to the flexible configuration of the approach, we may select any plane in the scene as the reference or X-Y plane. Assign \(I_1\) as the reference image, the first step is to find feature points to relate each image pair \((I_1, I_j), j = 2, 3, 4, ...M\). One suggested automatic approach for the procedure is a combination of Scale-Invariant Feature Transform (SIFT) [44] and Random Sample Consensus (RANSAC) [19].

SIFT is an algorithm in computer vision to detect and describe local features in images. The first step is to extract SIFT keypoints of objects from a set of images and store them in a database. By individually comparing each feature from a new image
to this database and finding candidate matching features based on Euclidean distance of their feature vectors, an object is recognized. Successful matches are filtered out from the full set as subsets of keypoints that agree on the object and its location, scale, and orientation in the new image. Consistent clusters are determined by using an efficient hash table implementation of the generalized Hough transform. Further detailed model verification is then applied to each cluster of 3 or more features that agree on an object and subsequently outliers are discarded. Finally, the probability is computed to test if a particular set of features indicates the presence of an object. Matches that pass all these tests can be considered correct with high confidence.

RANSAC is an iterative approach for the estimation of parameters in a mathematical model. For a set of observed data, a basic assumption is that it consists of "inliers" and "outliers" which indicate whether a distribution can be explained by some set of model parameters or not. Possible sources of the outliers include extreme values of noise, erroneous measurements or incorrect hypotheses about the interpretation of data. After n point correspondences are established, RANSAC can be applied
to provide robust estimation of the 2D homography. The proposed algorithm utilizing SIFT and RANSAC for finding corresponding planar points automatically is as follows:

**Automatic approach for finding planar point correspondences:**

1. Apply SIFT to extract and determine matching feature points.

2. Randomly choose 4 correspondences and check to avoid collinear points.

3. Assume that these feature points lie on a plane. Compute the homography \( H \) by normalized DLT from the 4 correspondences.

4. Apply RANSAC and find the \( H \) with most inliers. The images of points on the plane are assumed to be those which fit the approximate homography.

If the point set \( \{ X_n : n = 1, 2, ..., N \wedge N \geq 4 \} \) is found to be on the plane \( \pi_0 \) and its projections on images \( I_1 \) and \( I_j \) are \( \{ x_1 \} \) and \( \{ x_j \} \) respectively, the homography \( H_{1j} \) induced by plane \( \pi_0 \) such that \( x_1 = H_{1j}x_j \) can be estimated from SVD algorithm described in chapter 2. Note that even though a common point set provides convenience in computation, the feature points are not necessarily observed in all images since the homography is estimated for each image pair \( (I_1, I_j) \) separately.

As presumed in related literature on silhouette based volumetric recovery methods [18], we require silhouettes of objects are provided in all images (see Figure 4.1). Object silhouettes can either be extracted manually or can be generated in the sequel of a segmentation [17] or background subtraction [81] method. After establishing the correspondences for each image pair, all silhouettes are mapped onto a reference image by homography transformation. In [33], the cross-section of the object on the plane
is segmented using a minimizing energy functional under the level set framework. In the proposed approach, we rather follow a more intuitive voting procedure. Based on the fact that when all points in multiple images are warped onto a reference image by the homography transformation, theoretically only the points on the plane which induces the homography will coincide. Other points not on the plane shall show discrepancies as depicted in Figure 4.1. Hence, the cross-section with highest voting represents the object shape on this plane. In this setting, the mean of the mappings of the silhouettes from $I_j$ to the reference image using the homography transformations $I_{1j} = H_{1j}I_j$:

$$I_{\text{mean}} = \frac{1}{M} \left( I_1 + \sum_{j=2}^{M-1} I_{1j} \right)$$ (4.1)

generates an image of the object cross-sections which highlights the positions of the objects on that plane. This observation is illustrated in Figure 4.1, where all transformed silhouettes coincide at the regions corresponding the object cross-section where the hypothetical plane intersects with the object volume. Considering that silhouettes are binary masks, 1 indicating presence of the object, $I_{\text{mean}}$ will have pixels valued ranging from 0 to 1. From the characteristic of homography transformation, the areas of value 1 indicate locations in which an object cross-section exits. In context of shape recovery, we only consider the outline of the cross-section, which we refer to as the bounding-curve. Since the highlighted cross-section is usually thresholded to create a binary image, the bounding curve can be extracted by Canny or any other edge detector easily.

Following the discussion on subspaces in $\mathbb{P}^3$, a set of hypothetical planes generated from translating or rotating a reference plane constitute a subspace. A visualization of the parallel and rotated planes intersecting the same target object is demonstrated
in Figure 4.2. Depending on how the subspace is parameterized, each specific plane is determined by its vertical distance or angle to the reference plane. These parameters in the approach are implemented as $Z = \rho \Delta Z$ for parallel and $\theta = \rho \Delta \theta$ for rotated planes, in which $\rho$ is any integer, $\Delta Z$ and $\Delta \theta$ are the preset increments between adjacent planes. For each value of $Z$ or $\theta$, a new set of pixel correspondences is established from (3.4) and (3.21) respectively. These additional pixel correspondences are used to estimate homography transformations corresponding to the hypothetical planes, which are used to generate bounding-curves from (4.1). The stack of bounding-curves provides a representation of object in the image space, such as shown in Figures 4.3(a) and 4.3(b) for the two subspaces.
In order to retrieve the 3D Euclidean coordinated of the object, we need to back-project the bounding curves to the object space. The back-projection may be achieved in various ways. One approach is to use a set of feature points with known absolute object coordinates, such that the relation between the object space and the image space can be estimated. This can be realized by:

- Setting a specific feature such as a box,
- Using the features on a building with known dimensions,
- Using the relative length ratio between linear features.

Using one of these, one can generate a local Euclidean coordinate frame in the object space and the metric shape recovery can be achieved up to a scale. Theoretically,
one can pick up any measurable feature on the reference plane even though the axes are not orthogonal, but ideally a square is preferred for achieving a robust metric reconstruction.

In this research, a different approach is followed which eliminates the use of features that are known or extracted in the object space. A conjecture is brought up that the ground plane in the object space is identical to the reference image, such that the reference image is either affine rectified or acquired by an affine camera. In this setting, the shapes and coordinates of the objects on the ground plane appear exactly as those in the reference image; hence, all the bounding-curves, which are warped by respective homographies onto the affine rectified image, provides hypothetical planes at preset $Z$ or $\theta$ intervals. In the case of a perspective image of the ground plane which is conjectured to be acquired by an affine camera, the recovered shape will be projectively distorted.

If the metric recovery is required, before back-projecting the extracted bounding-curve to the object space, it is first mapped onto the ground image by the homography transformation estimated from $x'$ and $x$ for the parallel plane configuration, or from $x_\theta$ and $x'$ for the rotated plane. The transformed bounding-curve is then back-projected to the object space by the preset coordinates mapping, which is another homography. Thus, different from other approaches, the 3D shape recovery involved in the proposed approach consist of only 2D-2D transformations. While the X and Y coordinates of the 3D points are obtained directly from the back-projection, their Z coordinate is treated separately. In the parallel planes configuration, $Z$ is preset in equation 3.4 for each hypothetical plane, while in the rotated planes the $Z$ coordinates for each point varies with $Y$, and have to be reassigned as $Z = Y \tan \theta$. We should note that,
unless the absolute object coordinates, length ratios or line orthogonalities for the scene is available, the relation between the metric object space and the projective space cannot be established, and an unknown 3D homography between the recovered 3D and the Euclidean 3D exists. Hence, a Euclidean recovery can be obtained by choosing the reference image as an ortho-rectified or an ortho-view image. For affine-rectified reference image, the 3D recovery is up to an affine scale. We should also note that, when available, measured features are preferably used to achieve metric recovery.

The density of the recovered 3D points depends on $\Delta Z$ for parallel plane topology or $\Delta \theta$ for rotated plane topology. Apparently smaller increment creates larger number of planes slicing the object, hence provides better shape representation. In order to retrieve the complete recovery of object shape, the increments $\Delta Z$ or $\Delta \theta$ should be infinitesimal and there must exist infinite number of images covering every view of the object. Since these requirements are not feasible in practical cases, the shape recovered by the proposed approach can only be regarded as an approximation, nevertheless, it still provides invaluable 3D information for visualization of the scene. Another observation on the approach reveals that, due to the fact that these planes intersect all objects in the scene, the proposed approach is capable of recovering the shapes of multiple object simultaneously using a very simple to implement procedure.

4.2 Discussion

The proposed approach utilizes silhouettes and homography transformations to recover object shape. The detail of recovered shape is affected by the number of views as well as their distribution in the object space. Due to the intrinsic characteristics of
silhouettes, for small number of views, occluded object parts remain invisible which hinders recovering concavities and convexities. Nonetheless, for symmetric objects, as will be discussed in the experiments, one can use as few as two views for 3D recovery. In addition, for the images with missing observations, the threshold M in 4.1 can be reduced to accommodate the problem.

The major difference between the rotated and parallel plane topologies is that, the estimation of scale factors in the rotated case removes the requirement of estimating a vanishing line, but instead add the assumption that the distance from the reference feature to the rotation axis is known. However, both techniques induce similar limitations. During the recovery process, it is possible to observe singularities when the parameterized hypothetical planes approach the camera center in a particular view. For parallel plane topology an illustration is given in Figure 3.6. In this case, the mapping of the silhouette converges to a straight line preventing the extraction of bounding-curves for the scene objects, such that no shape information can be recovered. In order to avoid this problem, a practical solution is to have the reference image taken from a higher altitude or at a near-nadir angle to the reference plane.

There are also limitations related with certain subspace topologies. Particularly for the rotated planes, equation (3.24) has an degenerate configuration with its related geometry given in Figure 3.8. This degenerate condition arises at angle $\theta = \pi/2$ where all the estimated image pixels either lie at infinity or coincide with the vanishing point $v_z$. In the implementation of this research, these conditions are avoided by (1) testing $l_{axis}^T x_{\theta} \neq 0$ where $l_{axis}$ is the image of rotation axis, and (2) thresholding $|\theta| \leq \theta_{threshold}$. The purpose of the first test is to eliminate the situation in which estimated point set $\{x_{\theta}\}$ on $\pi_{\theta}$ all coincide with the image of rotation axis, meaning that $\pi_{\theta}$ is
perpendicular to the image plane. Due to the fact that $\tan \theta$ increases rapidly beyond $\theta = \pi/4$, for the second condition, more reliable results can be expected when the rotation axis is selected farther from all scene objects, such that $\theta < \pi/4$ is assured. Ultimately, if the rotation axis is chosen at the vanishing line, the rotated plane topology can be shown to reduce to the parallel plane topology.

Despite degenerate conditions outlined above, we should note that the proposed approach performs 3D recovery using simple linear transformations which are generated from as few as four corresponding coplanar points across the views and a vertical vanishing point in each image. These requirements are significantly lower than traditional shape recovery algorithms which exploit camera matrices and other impractical requirements. This observation provides implementation ease and significantly reduces the computational complexity.

4.3 Error analysis

![Diagram](image.png)

Figure 4.4: Error between the estimated and the true vanishing point.
The quality of recovered object shape is dependent on the robustness of pixel correspondences as well as the position of the vertical vanishing point. Under the condition that the pixels are sub-pixel accurate, or that there are many pixel correspondences, it can be shown that the mapping error between the images is negligible. In addition, the localization error of pixels can be considered to have a zero mean gaussian error. Since homography transformation is a linear mapping, the resulting scales and the homography transformations will also have zero mean gaussian errors. However, the vanishing point lies at infinity with zero scale and have a nonlinear effect on recovery process. Hence we consider, the main source of error in the proposed approach is attributed to the error in the vanishing point location. Estimated vanishing point \( \hat{v} \) can be expressed as the sum of true vanishing point \( v \) and an error vector \( \delta_v = r[\cos \alpha, \sin \alpha, 0]^\top \) where \( r \) is the pixel distance from \( v_z \) to \( \hat{v} \), and \( \alpha \) is the angle from x axis to the vector \( v_z \hat{v} \), as shown in Figure 4.4. Following this observation, we analyze the effect of localization error in the vanishing point on resulting mappings between images for both the parallel and rotated plane topologies. The analysis outlined in the following discussion is conducted using the two camera configurations provided in the Middlebury multi-view temple dataset [63]. Instead of using the silhouettes of the temple, we generated a synthetic dataset consisting discrete points in order to better observe the error in mappings with changes in \( r \) and \( \alpha \) parameters of the error vector \( \delta_v \).

### 4.3.1 Errors in parallel plane approach

In order to analyze the effect of error in localization of the vanishing point, two synthetic squares are placed in the object space at different heights. The first of
the squares is positioned on the ground plane, $Z = 0$, with four corners marked at $X_i^T = \{[0.1,0.1,0,1], [0.1,-0.1,0,1], [-0.1,-0.1,0,1], [-0.1,0.1,0,1]\}$. The second square is position on the plane at $Z = 0.2$ with four corners located at $X_i^T = \{[0.1,0.1,0.2,1], [0.1,-0.1,0.2,1], [-0.1,-0.1,0.2,1], [-0.1,0.1,0.2,1]\}$. For the sake of testing the accuracy of recovered points, we select a set of 3D points on plane $Z = 0.2$ at grids with 0.01m spacing in both X and Y direction. Using the camera projection matrices available in the dataset, we estimated subpixel accurate pixel locations $x_1 \ldots x_4, x'_1 \ldots x'_4$ corresponding to the corners of the squares. The projection of both the square and the grid pixels are shown in Figure 4.5. In this configuration, the distance from the true vanishing point $v_z$ to $x_i$ is about 2000 pixels in average. In order to examine the effect of error under some extreme conditions, we take 1/4 of this distance as $r$, which is significantly large and only possible when obvious distortion exists in the images.

In the proposed approach, the effect of vanishing point error is first manifested in the scale factor. This error can be observed by introducing the error vector $\hat{v}$ to
Figure 4.6: Errors on the scale factors under varying angle ($\alpha$) of the error vector when $r = 500$.

(3.5), which results in propagation of the error by:

$$\hat{s} = s + \delta_s = s + \frac{Z(x' - x)^\top \delta_v}{(x' - x)^T(x' - x)}$$

(4.2)

Figure 4.7: The errors on the estimated pixels increase with $Z$. 

59
From (4.2) and Figure 4.3.1, we observe that $\delta_s = 0$ when $\delta_v$ is perpendicular to the direction of $x'x$, and reaches the maximum when it is collinear with $x'x$. The pixel error also increases with the height $Z$ of the plane as a result from equation (3.4) and is shown in Figure 4.3.1.

![Figure 4.8: The distances from the estimated image pixels to the actual pixels at plane $Z = 0.2$ for different $r$ and $\alpha$ values.](image)

Considering the error vector $\delta_v$ of various combination of parameters $r$ and $\alpha$, when the scale factors and points $\{x'\}$ on the plane at $Z = 0.2$ are estimated accordingly, the maximum distances between the estimated points $\hat{x}'$ and their true locations are depicted in Figure 4.3.1. Here the magnitude of $r$ is extended to 1000, which is very large, to generate significant difference for the examination on the trend of error. Further inspection of Figure 4.3.1 also reveals that the maximum and minimum errors occur at circumstances contrary to the scale factor error. In other words, when the scale factor error is at its maximum, the pixel estimation error is at its minimum, and vice versa.
Figure 4.9: (a) The grid pixels transformed from another image using homography estimated from erroneous pixels. (b) Projection of the pixels in (a) onto the image of the ground plane.

Analysis above is about the effect of error within a single reference image $I_r$. The effects on shape recovery require other images containing significant localization errors in the vanishing points as well to be taken into consideration. For the purpose of analysis, here we adopt only one image $I_a$ and substitute silhouettes with grid points to demonstrate their effects. When the grid points are transferred to $I_r$ from $I_a$
using the homography estimated from erroneous pixels, apparent disparity between
the transferred grid pixels and the actual pixels at the plane $Z = 0.2$ is observed
in Figure 4.9(a). In the proposed approach, these pixels are further mapped to the
image of the ground plane as Figure 4.9(b) prior to the back-projection onto the
object space. As can be observed from the figures, if the object is bounded in the
cube, the recovered shape shall be qualitatively acceptable.

4.3.2 Errors in rotated plane approach

![Figure 4.10: Exact projection of the synthetic points.](image)

Similar to the analysis of error in parallel plane topology, points are selected on
the ground plane and on plane $\pi_{\theta}$ at angle $\theta = \pi/4$. The projection of these points
to the images are shown in Figure 4.10. Given the error in the vanishing point, the
scale factor computed from equation (3.23) results in propagation of the error by:

$$\hat{s}_z = s_z + \delta s_z = \frac{1}{t} \frac{\nabla_z (\nabla_v - x_{\theta})^\top (\nabla_v - x_{\theta}) + \delta_v (\nabla_v - x_{\theta}) + \delta_v (2(\nabla_z - x_{\theta}) + \delta_v)}{1}.$$  (4.3)
Figure 4.11: Errors on the scale factors for the rotated planes with a change in the angle (α) of the error vector for r = 500.

Except for the scale factor error in Figure 4.3.2, the behavior of the estimated pixels revealed from Figures 4.3.2, 4.3.2, 4.14(a) and 4.14(b) is similar to the errors observed for the parallel plane topology. Note that in both tests, the errors in the vanishing points are set to values relatively large, which rarely occur in practical cases, yet the disparity between the recovered and true shapes is not as significant as shown in Figures 4.9(b) and 4.14(b). The accuracy and completeness of the recovered 3D shape remains more affected by a set of other factors such as number of views and the quality of silhouettes which are beyond the scope of this analysis.

4.4 Summary

In order to manifest the complete procedure of the approach, a work flow provided in Figure 4.15 shows a set of steps required by the proposed 3D recovery approach.
The procedure is substantially similar for both parallel and rotated plane configurations except that in the scale factor estimation the vanishing line is not required for the rotated planes.

As is the situation in any state-of-the-art algorithm for shape recovery, the applicability of the approach is based on some requirements and assumptions, which is listed below.

**Silhouettes.** The silhouettes of the objects to be recovered need to be extracted prior to the image warping. As well as the number of views provided for the objects, the quality of silhouettes significant influence on the resulting shapes.

**Point correspondence.** Minimum of four point correspondences between each image and the reference image have to be identified. Note that these points may not be commonly observed in all images, since the required homographies are estimated from each image to the reference separately.
Figure 4.13: The distances from the estimated image pixels to the actual pixels at plane $\theta = \pi/4$ under different $r$ and $\alpha$.

**Vertical features.** Each image must contain one vertical feature with known or presumed height for computing the scale ratio $s$ of selected points. At least one pair of vertical ($Z$ direction) parallel lines are also required to retrieve the vertical vanishing point.

**Parallel lines.** Two parallel line pairs which are also parallel to the reference (ground) plane are used to estimate the vanishing line. However, if the vanishing line is approximated from the technique introduced in Section 3.2.2, this requirement may be substituted by additional vertical features. Note that in the approach using rotated planes, the vanishing line estimation is not performed, hence the requirement does not apply.

These requirements are not difficult to establish in real world environment, as will be demonstrated in the following chapter. Potentially they may be automatically obtained by combining other state-of-the-art algorithms, such as using SIFT.
Figure 4.14: (a) The grid pixels transformed from another image using homography estimated from erroneous pixels. (b) Projection of the pixels in (a) onto the image of the ground plane.

and RANSAC for establishing point correspondences. The simpleness of implementation and efficiency on computation result in a fast and practical approach for shape recovery.
Extract object silhouettes
Acquire multi-view images of a scene
Find point correspondences on the reference plane
Estimate homographies across images
Compute scale factor $s$ for each point
Estimate vanishing point and vanishing line (for parallel case only)
Change $Z$ or $\theta$ and compute new point set
Back-project bounding-curve to the object space
Warp silhouettes to the reference image and generate bounding-curve
No bounding-curve found
3D object shape
Stop

Figure 4.15: Work flow of the proposed approach.
CHAPTER 5

Experiments

In order to verify the validity of the proposed approach, five sets of experiments are conducted using image datasets with different characteristics such as symmetry and varying illumination. Due to the unavailability of standard datasets, except for the Middlebury multi-view dataset, we have taken images of objects ourselves and used images from public collections to perform qualitative and quantitative analysis. The required silhouettes for all the images are generated interactively using image segmentation tool GIMP®, and the novel views of recovered 3D points are created by MeshLab®.

5.1 Shape recovery for a single object

The first image set contains 12 images of a toy taken from different viewpoints (Figure 5.1(a)). For comparison reasons, both the parallel and rotated plane approaches are applied to the data set. The corners of the tiles (blue dots in Figure 5.1(a)) provide points for homography estimation and two of the points are selected to be lying on the axis of rotation (red line in Figure 5.1(a)). The two pens whose height are assumed to be 15cm are placed on the ground to provide linear features for the estimations of the vanishing point and the scale factors. The silhouettes of
the toy are shown in Figure 4.1. For the parallel plane topology, bounding-curves are generated for the planes with 1cm increment. After back-projecting the stack of bounding-curves to the object space, the shape of the toy is approximated by over 44,000 3D points. However, due to few number of views, some fractals appear around the waist of the toy (Figure 5.1(b)).
For the rotated plane topology, bounding-curves are generated using 8 images for planes at angle $\theta$ ranging from $0^\circ$ to $45^\circ$ with $1^\circ$ increment. The back-projected bounding-curves create over 55,000 3D points. The detailed shape such as the toes can be observed from the reconstructed shape. Compare to the visual result presented in Figures 5.1(b), the approach creates smoother shape using even fewer images. This can be attributed to eliminating the estimation of additional vanishing line. Extra post processing can be used to further improve both resulting 3D surfaces but is beyond the context of this dissertation.

In both the experiments, the processing time for generating ten thousands of points using an coarse Matlab implementation is within an hour, and varies with the image size and the number of images used. Though compared to other prevailing photo-consistency based approaches, the reconstructed shape is not as smooth, for the cases where the constraints of other methods are not met, the proposed approach provides an alternative way for the approximation of object shapes.

### 5.2 Human body recovery

In this section, two sets of images are used to test the shape recovery for human bodies. The first set contains seven low quality images of an individual in the scene (Courtesy of Dr. Ajit Chaudhari, the Ohio State University). In this set, uncalibrated images were taken by several wide-angle lens cameras simultaneously and contains significant radial distortion. The distortion is removed, though not rigorously, using the fact that some objects in the scene contain straight edges. The top view shown in Figure 5.2(g) is selected as the reference image and the object space coordinates is set to be identical with this image. Due to the absence of common features identifiable in
Figure 5.2: Test images for human shape recovery using parallel planes.
Figure 5.3: Recovered shape for a human body using parallel planes.

all images, correspondences are evaluated separately from each image to the reference. An important quality problem that is observed from some of the original images that the illumination is dim and some parts of the body can not be well distinguished from the background, hence results in uncertainties during the silhouette extraction. In order to overcome this problem, we selected a bigger and approximate region for the obscure parts, such as the left leg in Figure 5.2(c). After the homography mappings from all images to the reference image, the extra regions are filtered out due to the fact that these regions are not present in all images. The recovered shape revealed in Figure 5.3 shows that the outline of the torso and limbs is well presented despite the noisy silhouettes.
Figure 5.4: Shape recovery for human body using rotated planes.

The second image set for testing the rotated plane approach contains six views of a female standing in the scene, two of which are shown in Figure 5.4. Note that the locations of the box are different in these views. The box functions as the required vertical feature to provide vanishing point as well as the reference scale factor. According to the setting for the rotated plane topology, as long as the height of the feature as well as its distance to the rotation axis are gaugeable, it can be positioned anywhere in the scene. The bounding-curves are generated for angles ranging from 0° to 45° with 1° increment and from 45° to 75° with 0.5° increment. Due to the fact that tan θ increases rapidly, consequently creates larger intervals between planes, the incremental angle is reduced beyond 45° to avoid sparseness of the points. The
recovered 3D shapes is shown in Figure 5.5. As can be observed from the figure, the shape below the waist is smoother compared to the head.

In both the experiments, the processing time for generating ten thousands of points using an unoptimized Matlab implementation is within an hour, and varies with the image size and the number of images used. The accuracy of reconstructed 3D models is only qualitatively evaluated due to missing ground truth.

5.3 Symmetric object shape

The third experiment is conducted on an image set with only two views of the Taj Mahal (India), which are shown in Figure 5.6(a) and 5.6(b). The images are selected from the search result provided by Google© and have different spatial resolutions.
Figure 5.6: Two images of Taj Mahal, India, selected and downloaded from the search result provided by Google®.

For this experiment due to limited views, we selected four pixel correspondences as the top of four detached minarets springing from the plinth of the Taj Mahal. Based on the knowledge that the building is symmetric, each image is assumed taken from front, rear, right and left views of the building, equivalently providing three additional views for the shape recovery. Furthermore, by incorporating the fact that the four selected pixels form a square in the object space, a local object coordinate system
Figure 5.7: Recovered shape of Taj Mahal

can be established accordingly. Recovered 3D shape of the building is shown in Figures 5.7(a) and 5.7(b). Since we did not use metric information, such the height
or distances between the points, the recovered shape is only up to a metric scale. The result suggests that, for symmetric objects, synthetic views can be created and the number of required images for recovering the shape can be reduced significantly.

5.4 Building shape and comparison with LiDAR data

The experiment is designed to perform quantitative analysis of the proposed approach. In order to facilitate this, we selected a target building located in downtown Columbus, Ohio (Figure 5.8(a) and 5.8(b)) for which the LIDAR (Light Detection and Ranging) data is obtained through the Ohio Statewide Imagery Program (OSIP). The LIDAR point cloud is interpolated to grids with spacing of 4m × 4m (Figure 5.10(a)). According to the LIDAR data, the height from the roof to the ground is about 240m. The points with elevations ranging from 975.82m to 1007.67m are inspected to be on or above the roof. For corresponding image set, we obtained eight oblique images (bird’s eye view) from Microsoft VirtualEarth© in addition to a top view provided by Google Maps©. After recovering the 3D building model using the proposed approach, a comparison is performed on the quality of the recovered scene to the LiDAR points which is used as the “ground truth”. Particularly, two quality indices proposed in [64] are adopted, which include:

- **accuracy**, which describe how close the recovered model is to the ground truth model, and
- **completeness**, which indicate how much ground truth is modeled by the recovered model.

The accuracy represented by the mean of Euclidean distances between corresponding grid points in two datasets, is 5m. The completeness reaches 89% which is the
percentage of LiDAR grid points that have been recovered by the 3D model. Qualitatively, as shown in Figure 5.4, the recovered roof is not flat due to the lack of a side-view. This results in ambiguous cross-sections that exist above the roof. Hence, the two structures on top of the roof are not well delineated. However, the recovered
model provides smoother outlines of the building. More specifically, compared to the LIDAR data which has many missing 3D points on the facades of the building, the proposed method recovered a complete 3D model.

5.5 Comparisons with other approaches

The last experiment is conducted on one of the image sets provided in the Middlebury Multiple-view dataset [63]. The data set includes two sets of images specifically generated for comparison of volumetric shape recovery methods based on photo consistency. We particularly used the Temple image set, which contains 312 images of a plaster reproduction of “Temple of the Dioskouroi” in Agrigento, Sicily. The images were captured from various viewpoints covering about 80% of a hemisphere surrounding the object using a CCD camera with a resolution of 640 × 480 pixels. A 3D model is generated by a laser stripe scanner and is taken as a reference for the evaluation and comparison of proposed approach. The vanishing points required by
Figure 5.10: Elevation grids created from LiDAR data and the recovered 3D model.

the proposed approach are obtained directly from the third rows of the projection matrices for corresponding images, hence considered error-free. The silhouettes are created by semi-automatically subtracting the dark background in the images, which
Figure 5.11: Two images from the temple dataset provided by [63].

resulted in object boundaries that affected the recovery result. The model generated from the approach achieves 1.57mm accuracy and 80% of completeness using the evaluation scheme described in [64]. We should note that the proposed method relies on object silhouettes, which hides shape, such as concavities, details otherwise available from shading information. These details usually manifest themselves as concavities and convexities in the 3D object shape. This fact results in a lower accuracy when compared against photo consistency based methods whose performances are listed at http://vision.middlebury.edu/mview/eval/. A tabulated quantitative performances of proposed approach is listed in Table 5.5.
Figure 5.12: Reconstructed temple models. (a) The dense 3D point cloud generated from the proposed approach. (b) 3D model from laser scanning provided by [63].

<table>
<thead>
<tr>
<th>Object</th>
<th>Known object dimension</th>
<th>Accuracy</th>
<th>Completeness</th>
<th>Error in percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temple</td>
<td>10cm × 16cm × 8cm</td>
<td>1.57mm</td>
<td>80%</td>
<td>0.9% - 1.9%</td>
</tr>
<tr>
<td>Building</td>
<td>240m (height)</td>
<td>5m</td>
<td>89%</td>
<td>2.1%</td>
</tr>
</tbody>
</table>

5.6 Summary

In this chapter the experiments are designed to verify the capability of the approach. It has been seen that reasonable results are obtained even using only few images. From examining the experimental results, one may argue that the generated
models from proposed approach are far from perfect. Further improvements, including visual effects and geometric accuracy, are possibly accomplished by incorporating other techniques in the future.
CHAPTER 6

Conclusion and future work

In this dissertation we have presented a new 3D scene recovery approach by introducing parameterized hypothetical planes in $\mathbb{P}^3$ and analyzing the projective geometry between images with respect to these planes. It is shown that different parameterizations of the hypothetical planes provide different subspace topologies which significantly reduces the computational complexity of 3D recovery. Unlike traditional recovery algorithms, the proposed approach generates coarse approximation of object shape even when camera parameters and poses are not available. In addition, it does not require fundamental matrix estimation, visual hull generation, or dense point correspondences across images. The implementation of the approach is relatively simple and generates 3D shapes very fast. The experimental results demonstrate that our method is capable of recovering object shape in the scene efficiently by using as few as two views for symmetric objects and a minimum of four point correspondences across images.

In summary, the merits of the proposed approach include:

1. **Efficiency**: The proposed method does not require camera calibration and estimation of the fundamental matrix; hence, the computational complexity
related to establishing abundant number of point correspondences is reduced. The object shape is reconstructed using the apparent contours of the projected silhouettes, which provide the surface of the 3D object without the necessity of estimating visual hulls.

2. **Flexibility:** The level of detail in the reconstructed object is dependent on the number of images used, as well as the distance or angle between the planes, which provide a balance between the computation time and the smoothness of the recovered shape. Furthermore, since no dense point correspondences are needed and the missing information can be recovered from other images of different views, the use of object silhouettes automatically eliminates the problems related to occlusion.

3. **Practicability:** The set of simple equations involved in our approach are possibly computed in real time, providing fast metric recovery of the complete scene for use in different problem domains such as scene understanding and object tracking. By using techniques for finding points on the reference plane [44], and generating the silhouettes [71], the proposed approach achieves automated reconstruction. In addition, the 3D shapes of all the objects in the scene are recovered simultaneously as the planes cuts all the object volumes in the object space.

Although the proposed approach was applied successfully to the data sets in the experiments, some issues still need to be overcome, such as:
1. The concavity on the object surface can not be modeled by the approach since the information is hidden in silhouettes. This is taken as the intrinsic limitation for silhouette-based approaches and so far there is no solution to this problem.

2. Some of the 3D points are redundant, since it is possible that several image points are mapped to the same 3D location. These points should be removed for faster 3D model generation.

3. Currently the recovered shapes are represented with 3D point clouds. Other representations such as grid or triangulated mesh may be generated by incorporating advanced surface modeling techniques to provide more photo-realistic views of the objects.

These concerns in the existing approach will be addressed in the future to achieve better performance. Several potential research directions which may as well provide solutions to an improved recovery are proposed, including:

1. **Removal of vertical vanishing point estimation.** In this approach, the vertical vanishing point serves as the key to recover 3D shape from 2D images. Although the provided analysis shows that the errors introduced by the localization of vanishing points have limited impact on the recovered shape, removal of the vanishing point in the procedure can further improve the accuracy of recovery. In addition, the estimation of vanishing points requires parallel linear features in the scene, which may not be available in some cases. Theoretically any point which lies on any plane other than the reference plane can also provide information about the vertical direction, hence, an advanced approach may exclude the vanishing point but use other point correspondences instead.
2. **Integration with color information.** Inability of the approach to model concavity can be solved when color information is taken into consideration. The procedure usually requires visibility test, which may increase computational complexity thus has not been realized in the approach at current stage. The possibility of integrating photo-consistency based algorithms with the proposed approach will be studied in order to achieve higher recovery accuracy.

3. **Scene recovery.** The proposed approach focuses on recovering specific objects in the scene. Since the hypothetical planes slices not only the interest objects but any point in the scene, the recovery may be extended to the entire scene to generate even more realistic views.

4. **Feature-based photogrammetry.** Conventional photogrammetric procedures such as aerial triangulation are based on point-based approaches. Recently, researchers turn to feature-based approaches in the attempt of making photogrammetric processes more robust. According to the characteristics of homography transformation, the homography induced by a plane can also be estimated from corresponding lines. The capability of mapping line features by a simple linear transformation may benefit the development of feature-based photogrammetry.

5. **Other representations of homography** According to the scene and camera configuration, homography induced by a plane may have a specific form, such as affine homography and infinite homography. These special homographies are currently adopted separately by different approaches. A new representation of mapping incorporating the combination of specific homographies which may
help eliminate some of the requirements is under development. The effectiveness
of the new approach will be evaluated.

In the future work, extensive study will be placed on addressing the aforemen-
tioned concerns and research possibilities, and finally a more adaptable recovery ap-
proach shall be realized.
BIBLIOGRAPHY


