Chapter 1

Introduction

Methods to digitize and reconstruct the shapes of complex three dimensional objects have evolved rapidly in recent years. The speed and accuracy of digitizing technologies owe much to advances in the areas of physics and electrical engineering, including the development of lasers, CCD's, and high speed sampling and timing circuitry. Such technologies allow us to take detailed shape measurements with precision better than 1 part per 1000 at rates exceeding 10,000 samples per second. To capture the complete shape of an object, many thousands, sometimes millions of samples must be acquired. The resulting mass of data requires algorithms that can efficiently and reliably generate computer models from these samples.

In this thesis, we address methods of both digitizing and reconstructing the shapes of complex objects. The first part of the thesis is concerned with a popular range scanning method known as optical triangulation. We show that traditional approaches to optical triangulation have fundamental limitations that can be overcome with a novel method called *spacetime analysis*. The second part of this thesis concerns reconstruction of surfaces from range data. Many rangefinders, including optical triangulation scanners, can acquire regular, dense samplings called *range images*. We describe a new method for building complex models from range images in a way that satisfies a number of desirable properties.

In this chapter, we describe some of the applications of 3D shape acquisition (Section 1.1) followed by a description of the variety of acquisition methods (Section 1.2). In

Section 1.3, we pose the problem of surface reconstruction from range images. In Section 1.4, we place the goals of this thesis in the context of Stanford's 3D Fax Project. In Section 1.5, we describe the contributions of this thesis, and in the last section we outline the remainder of the thesis.

1.1 Applications

The applications of 3D shape digitization and reconstruction are wide-ranging and include manufacturing, virtual simulation, scientific exploration, medicine, and consumer marketing.

1.1.1 Reverse engineering

Many manufacturable parts are currently designed with Computer Aided Design (CAD) software. However, in some instances, a mechanical part exists and belongs to a working system but has no computer model needed to regenerate the part. This is frequently the case for machines currently in service that were designed before the advent of computers and CAD systems, as well as for parts that were hand-tuned to fit into existing machinery. If such a part breaks, and neither spare parts nor casting molds exist, then it may be possible to remove a part from a working system and digitize it precisely for re-manufacture.

1.1.2 Collaborative design

While CAD tools can be helpful in designing parts, in some cases the most intuitive design method is physical interaction with the model. This is especially true when the model must have esthetic appeal, such as the exteriors of consumer products ranging from perfume bottles to automobiles. Frequently, companies employ sculptors to design these models in a medium such as clay. Once the sculpture is ready, it may be digitized and reconstructed on a computer. The computer model is then suitable for dissemination to local engineers or remote clients for careful review, or it may serve as a starting point for constructing a CAD model suitable for manufacture.

1.1.3 Inspection

After a manufacturer has created a computer model for a part either by shape digitization of a physical model or through interactive CAD design, he has a variety of options for creating this part, both as a working prototype and as a starting point for mass manufacture. Ultimately, the dimensions of the final manufactured part must fall within some tolerances of the original computer model. In this case, shape digitization can aid in determining where and to what extent the computer model and the shape of the actual part differ. These differences can serve as a guide for modifying the manufacturing process until the part is acceptable.

1.1.4 Special effects, games, and virtual worlds

Synthetic imagery is playing an increasingly prominent role in creating special effects for cinema. In addition, video games and gaming hardware are moving steadily toward interactive 3D graphics. Virtual reality as a means of simulating worlds of experience is also growing in popularity. All of these applications require 3D models that may be taken from real life or from sculptures created by artists. Digitizing the shapes of physical models will be essential to populating these synthetic environments.

1.1.5 Dissemination of museum artifacts

Museum artifacts represent one-of-a-kind objects that attract the interest of scientists and lay people world-wide. Traditionally, to visualize these objects, it has been necessary to visit potentially distant museums or obtain non-interactive images or video sequences. By digitizing these parts, museum curators can make them available for interactive visualization. For scientists, computer models afford the opportunity to study and measure artifacts remotely using powerful computer tools.

1.1.6 Medicine

Applications of shape digitization in medicine are wide ranging as well. Prosthetics can be custom designed when the dimensions of the patient are known to high precision. Plastic surgeons can use the shape of an individual's face to model tissue scarring processes and visualize the outcomes of surgery. When performing radiation treatment, a model of the patient's shape can help guide the doctor in directing the radiation accurately.

1.1.7 Home shopping

As the World Wide Web provides a backbone for interaction over the Internet, commercial vendors are taking advantage of the ability to market products through this medium. By making 3D models of their products available over the Web, vendors can allow the customer to explore their products interactively. Standards for disseminating 3D models over the web are already underway (e.g., the Virtual Reality Modeling Language (VRML)).

1.2 Methods for 3D Digitization

A vast number of shape acquisition methods have evolved over the last century. These methods follow two primary directions: passive and active sensing. Passive approaches do not interact with the object, whereas active methods make contact with the object or project some kind of energy onto it. The computer vision research community is largely focused on passive methods that extract shape from one or more digitized images. Computer vision approaches include shape-from-shading for single images, stereo triangulation for pairs of images, and optical flow and factorization methods for video streams. While these methods require very little special purpose hardware, they typically do not yield dense and highly accurate digitizations required of a number of applications.

The remainder of this section summarizes the active range acquisition methods. Figure 1.1 introduces a taxonomy which we follow. This taxonomy is by no means comprehensive; rather, it is intended to introduce the reader to the variety of methods available.

Among active sensing methods, we can discern two different approaches: contact and non-contact sensors. Contact sensors are typically touch probes that consist of jointed arms or pulley-mounted tethers attached to a narrow pointer. The angles of the arms or the lengths of the tethers indicate the location of the pointer at all times. By touching the pointer to the surface of the object, a contact event is signaled and the position of the pointer is recorded. Touch probes come in a wide range of accuracies as well as costs. Coordinate Measuring

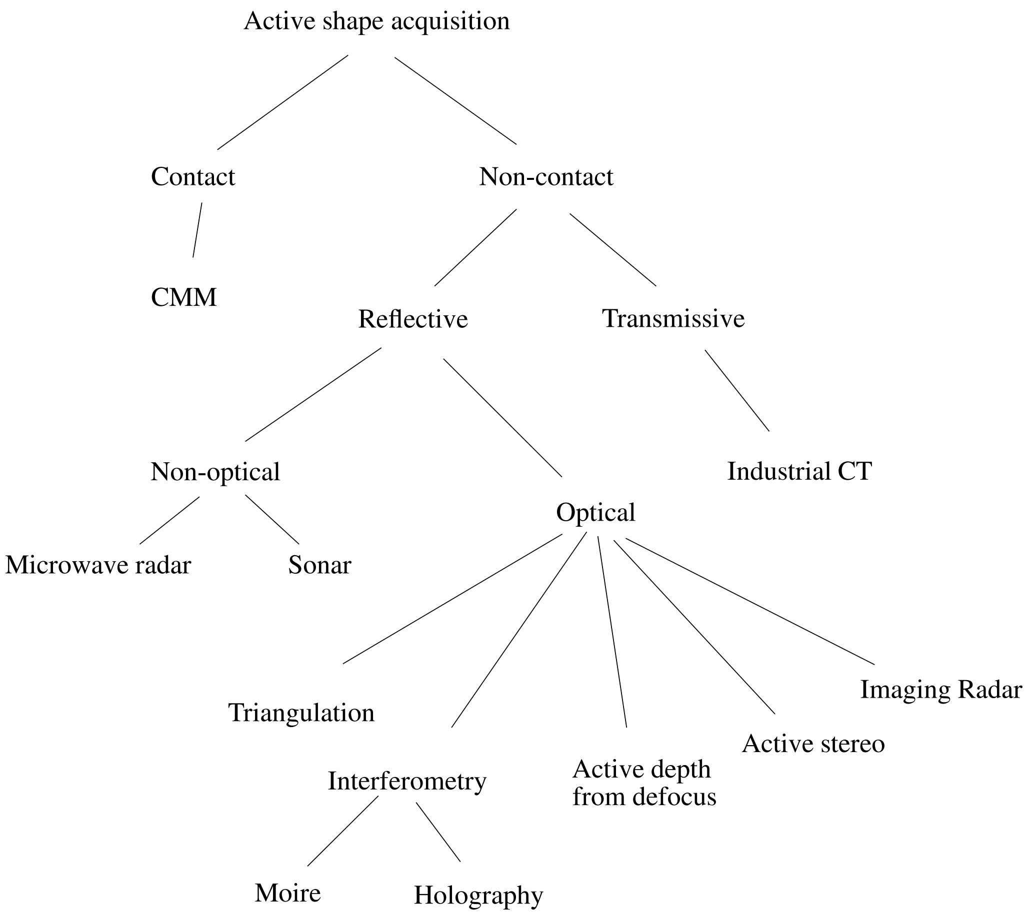

Figure 1.1: Taxonomy of active range acquisition methods.

Machines (CMM's) are extremely precise (and costly), and they are currently the standard tool for making precision shape measurements in industrial manufacturing. The main drawbacks of touch probes are:

- They are slow.
- They can be clumsy to manipulate.
- They usually require a human operator.
- They must make contact with the surface, which may be undesirable for fragile objects.

Active, non-contact methods generally operate by projecting energy waves onto an object followed by recording the transmitted or reflected energy. A powerful transmissive approach for shape capture is industrial computer tomography (CT). Industrial CT entails bombarding an object with high energy x-rays and measuring the amount of radiation that passes through the object along various lines of sight. After back projection or Fourier projection slice reconstruction, the result is a high resolution volumetric description of the density of space in and around the object. This volume is suitable for direct visualization or surface reconstruction. The principal advantages of this method over reflective methods are: it is largely insensitive to the reflective properties of the surface, and it can capture the internal cavities of an object that are not visible from the outside. The principal disadvantages of industrial CT scanners are:

- They are very expensive.
- Large variations in material densities (e.g., wood glued to metal) can degrade accuracy.
- They are potentially very hazardous due to the use of radioactive materials.

Among active, reflection methods for shape acquisition, we subdivide into two more categories: non-optical and optical approaches. Non-optical approaches include sonar and microwave radar (RAdio Detecting And Ranging), which typically measure distances to objects by measuring the time required for a pulse of sound or microwave energy to bounce back from an object. Amplitude or frequency modulated continuous energy waves can also be used in conjunction with phase or frequency shift detectors. Sonar range sensors are typically inexpensive, but they are also not very accurate and do not have high acquisition speeds. Microwave radar is typically intended for use with long range remote sensing, though close range optical radar is feasible, as described below.

The last category in our taxonomy consists of active, optical reflection methods. For these methods, light is projected onto an object in a structured manner, and by measuring the reflections from the object, we can determine shape. In contrast to passive and non-optical methods, many active optical rangefinders can rapidly acquire dense, highly accurate range samplings. In addition, they are safer and less expensive than industrial CT, with the limitation that they can only acquire the optically visible portions of the surface. Several surveys of optical rangefinding methods have appeared in the literature; the survey in [Besl 1989] is especially comprehensive. These optical methods include imaging radar, interferometry, active depth-from-defocus, active stereo, and triangulation.

Imaging radar is the same as microwave radar operating at optical frequencies. For large objects, a variety of imaging radars have been demonstrated to give excellent results. For smaller objects, on the order of one meter in size, attaining 1 part per 1000 accuracy with time-of-flight radar requires very high speed timing circuitry, because the time differences to be detected are in the femtosecond (10^{-12} second) range. A few amplitude and frequency modulated radars have shown promise for close range distance measurements.

Interferometric methods operate by projecting a spatially or temporally varying periodic pattern onto a surface, followed by mixing the reflected light with a reference pattern. The reference pattern demodulates the signal to reveal the variations in surface geometry. Moire interferometry involves the projection of coarse, spatially varying light patterns onto the object, whereas holographic interferometry typically relies on mixing coherent illumination with different wave vectors. Moire methods can have phase discrimination problems when the surface does not exhibit smooth shape variations. This difficulty usually places a limit on the maximum slope the surface can have to avoid ranging errors. Holographic methods typically yield range accuracy of a fraction of the light wavelength over microscopic fields of view.

Active depth from focus operates on the principal that the image of an object is blurred by an amount proportional to the distance between points on the object and the in-focus object plane. The amount of blur varies across the image plane in relation to the depths of the imaged points. This method has evolved as both a passive and an active sensing strategy. In the passive case, variations in surface reflectance (also called surface texture) are used to determine the amount of blurring. Thus, the object must have surface texture covering the whole surface in order to extract shape. Further, the quality of the shape extraction depends on the sharpness of surface texture. Active methods avoid these limitations by projecting a pattern of light (e.g., a checkerboard grid) onto the object. Most prior work in active depth from focus has yielded moderate accuracy (up to one part per 400 over the field of view [Nayar et al. 1995]).

Active stereo uses two or more cameras to observe features on an object. If the same feature is observed by two cameras, then the two lines of sight passing through the feature point on each camera's image plane will intersect at a point on the object. As in the depth from defocus method, this approach has been explored as both a passive and an active sensing strategy. Again, the active method operates by projecting light onto the object to avoid difficulties in discerning surface texture.

Optical triangulation is one of the most popular optical rangefinding approaches. Figure 1.2a shows a typical system configuration in two dimensions. The location of the center of the reflected light pulse imaged on the sensor corresponds to a line of sight that intersects the illuminant in exactly one point, yielding a depth value. The shape of the object is acquired by translating or rotating the object through the beam or by scanning the beam across the object. Due to the finite width of the light beam, inaccuracies arise when the surface exhibits significant changes in reflectance and shape. The first part of this thesis is concerned with describing the range accuracy limitations inherent in traditional methods, followed by the introduction of a new method of optical triangulation, called spacetime analysis, which substantially removes these limitations.

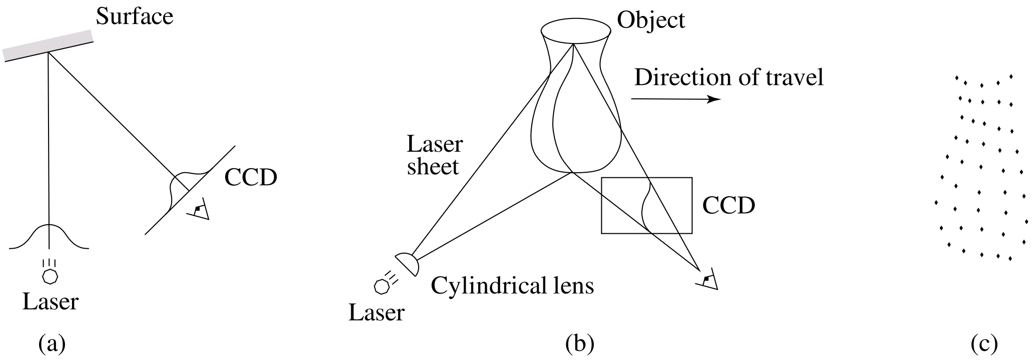

Figure 1.2: Optical triangulation and range imaging. (a) In 2D, a narrow laser beam illuminates a surface, and a linear sensor images the reflection from an object. The center of the image pulse maps to the center of the laser, yielding a range value. (b) In 3D, a laser stripe triangulation scanner first spreads the laser beam into a sheet of light with a cylindrical lens. The CCD observes the reflected stripe from which a depth profile is computed. The object sweeps through the field of view, yielding a range image. Other scanner configurations rotate the object to obtain a cylindrical scan or sweep a laser beam or stripe over a stationary object. (c) A range image obtained from the scanner in (b) is a collection of points with regular spacing.

1.3 Surface reconstruction from range images

In this section, we describe the goal of the second part of this thesis: surface reconstruction from range images. We begin with a brief explanation of range imaging.

1.3.1 Range images

Many active optical rangefinders sample the shape of an object along a set of regularly spaced lines of sight, yielding a grid of depths known as a range image. These rangefinders may be thought of as range cameras; i.e., they take pictures of a scene, but each image pixel contains a depth instead of a color. While conventional cameras typically take pictures that correspond to a perspective projection, the imaging geometry of rangefinders may vary widely, as discussed in Chapter 5. Figure 1.2 illustrates how a laser stripe triangulation scanner can acquire a range image. If we think of a surface as being described by a function, f, then a range image, $\hat{f}$ is a sampling of this surface, where each sample $\hat{f}(j, k)$ is the observed distance to the surface as seen along the line of sight indexed by (j, k).

1.3.2 Surface reconstruction

Each range image provides a detailed description of an object as seen from one point of view. Before attempting to reconstruct the entire shape of an object, we require multiple range images. As depicted in Figure 1.3, a single range image generally cannot acquire all sides of an object, thus requiring multiple range images to be taken. In fact, many objects are too complex to be captured by a small number of range images taken, for instance, from mutually orthogonal directions and their opposites (i.e., from the six faces of a cube). Due to self occlusions, an object may require a large number of range scans to see every visible point[1]. In some cases, points on the surface of an object are shiny enough to reflect most light that strikes them in a mirror-like fashion. Such points may only be seen when observed from particular angles, requiring multiple range images in order to see all of them.

Taking many range images offers two additional advantages: noise reduction and improved sampling rate. Real sensors suffer from noise due to a variety of causes including laser speckle and CCD scanline jitter. Multiple noisy sightings of a point on a surface can contribute to a reduced-variance estimate of the location of that point. Also, portions of a surface observed at a grazing angle from one point of view tend to be undersampled. Taking many range images increases the likelihood of sampling the surface when it is directly facing the sensor.

Since we require multiple range images to capture the shape of an object, we must devise a method to reconstruct a single description of this shape. The problem of reconstructing a surface from range images can be stated as follows:

> Given a set of p aligned, noisy range images, $\hat{f}_1, \cdots, \hat{f}_p$, find the 2D manifold that most closely approximates the points contained in the range images.

Note that in order to merge a set of range images into a single description of an object, it is necessary to place them all in the same coordinate system; i.e., they must be registered or aligned with respect to each other. The alignment may arise from prior knowledge of the pose of the rangefinder when acquiring the range image, or it may be computed using one

[1] Indeed, for some scanning technologies, points on the visible surface of an object may be "unacquirable" due to the geometry of the sensor. Optical triangulation scanners, for example, cannot acquire concavities inaccessible to a triangular probe with probe angle defined by the optical triangulation angle.

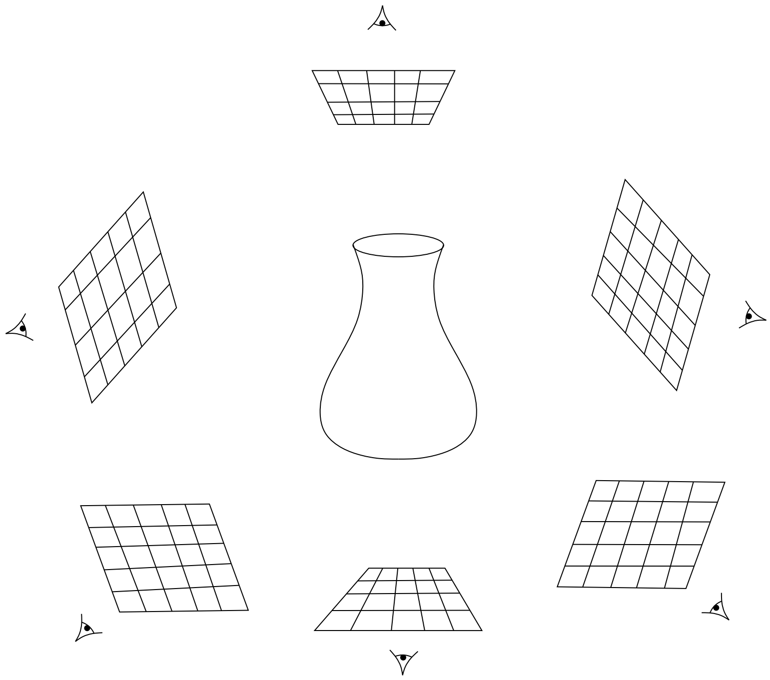

Figure 1.3: Range images taken from multiple viewpoints

of a number of algorithms. We assume in this thesis that the range images are aligned to a high degree of accuracy before we attempt to merge them.

A number of algorithms have been proposed for surface reconstruction from range images. Our experience with these algorithms has led us to develop a list of desirable properties:

- *Representation of range uncertainty*. The data in range images typically have asymmetric error distributions with primary directions along sensor lines of sight. The method of surface reconstruction should reflect this fact.

- *Utilization of all range data*, including redundant observations of each object surface. If properly used, this redundancy will provide a reduction of sensor noise.

- *Incremental and order independent updating*. Incremental updates allow us to obtain a reconstruction after each scan or small set of scans and allow us to choose the next best orientation for scanning. Order independence is desirable to ensure that results are not biased by earlier scans. Together, incremental and order independent updating allows for straightforward parallelization.

- *Time and space efficiency*. Complex objects may require many range images in order to build a detailed model. The range images and the model must be represented efficiently and processed quickly to make the algorithm practical.

- *Robustness*. Outliers and systematic range distortions can create challenging situations for reconstruction algorithms. A robust algorithm needs to handle these situations without catastrophic failures such as holes in surfaces and self-intersecting surfaces.

- *No restrictions on topological type*. The algorithm should not assume that the object is of a particular genus. Simplifying assumptions such as "the object is homeomorphic to a sphere" yield useful results in only a restricted class of problems.

- *Ability to fill holes in the reconstruction*. Given a set of range images that do not completely cover the object, the surface reconstruction will necessarily be incomplete. For some objects, no amount of scanning would completely cover the object, because some surfaces may be inaccessible to the sensor. In these cases, we desire an algorithm that can automatically fill these holes with plausible surfaces, yielding a model that is both "watertight" and esthetically pleasing.

Prior algorithms possess subsets of these properties. In this thesis, we present an efficient volumetric method for merging range images that possesses all of these properties.

1.4 The 3D Fax Project

The work described in this thesis fits into a framework for digitizing the shape and appearance of objects: the 3D Fax Project at Stanford University. Figure 1.4 shows the flow of data in this framework from scanning through construction of CAD models and manufacture of a hardcopy.

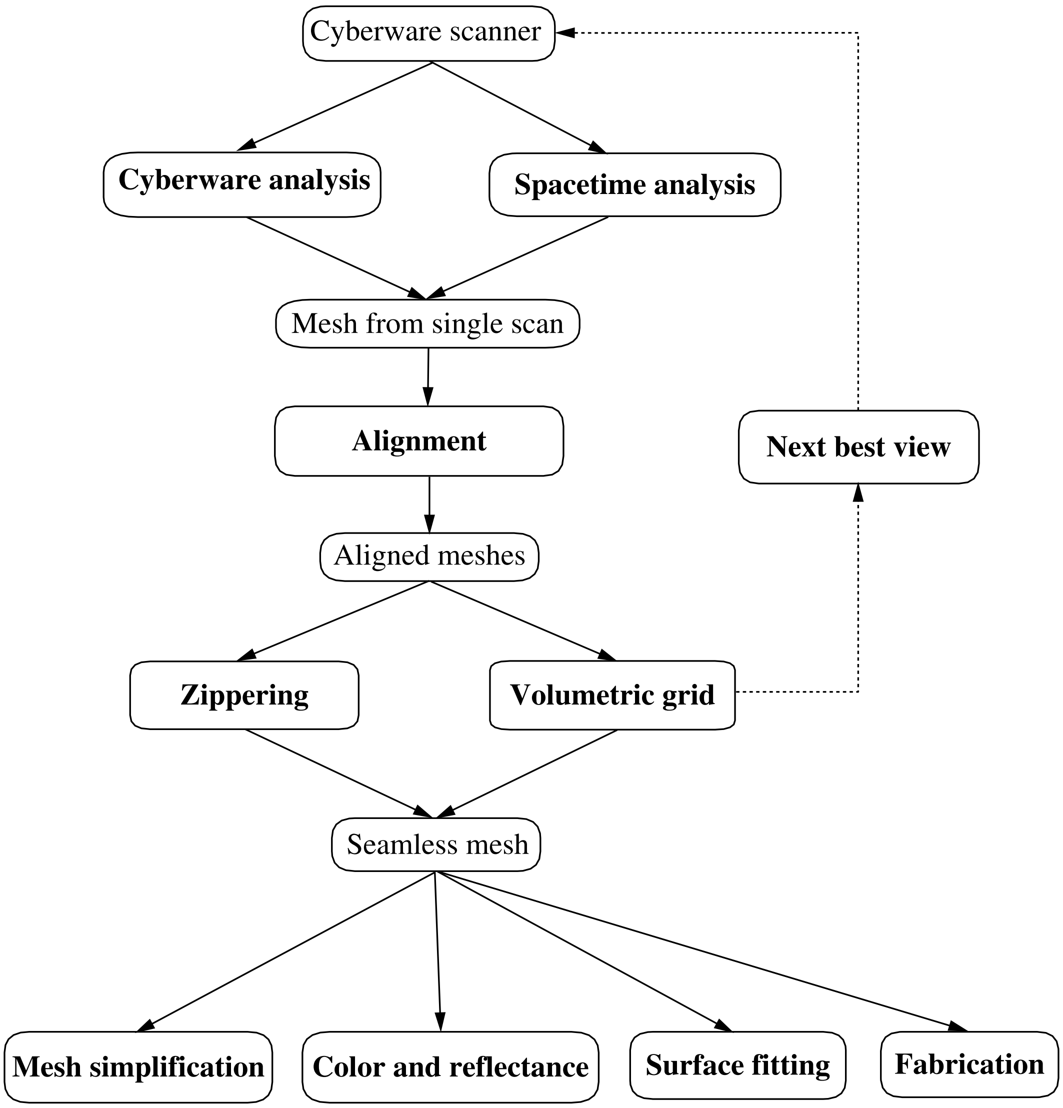

Figure 1.4: Data flow in the 3D Fax Project

Figure 1.5: A photograph of the Cyberware Model MS 3030 optical triangulation scanner used for acquiring range images in this thesis.

The input device for this project is an optical triangulation scanner manufactured by Cyberware Laboratories (pictured in Figure 1.5). The scanner performs triangulation in hardware using a traditional method of analysis which is prone to range errors. Alternatively, we can apply a triangulation method called "spacetime analysis" to minimize these range artifacts, as described in this thesis.

The range data obtained from the scanner is in the form of a dense range image. By repeatedly scanning an object from different points of view, we obtain a set of range images that are not necessarily aligned with respect to one another. Using an iterated closest point algorithm developed in [Besl & McKay 1992] and modified for aligning partial shape digitizations in [Turk & Levoy 1994], we transform the scans into a common coordinate system.

After alignment, the range images provide the starting point for performing a surface reconstruction. One method for merging the range images into a single surface is called "zippering." This method operates directly on triangle meshes, and while it is well-behaved for relatively smooth surfaces, it has been show to fail in regions of high curvature. An alternative method is to merge the meshes in a volumetric domain followed by an isosurface extraction. This method has a number of advantages over the zippering approach, as will be shown in this thesis.

After merging a set of range images, it is typically the case for a complex object that some portion of the object remains to be digitized. The problem of deciding how best to orient the object relative to the scanner is the so-called "next best view" problem, an active area of research in the 3D Fax Project. We believe that defining the emptiness of space around an object using the volumetric method described in this thesis will play an important role in developing an effective next best view algorithm.

Once a satisfactory model is obtained, it is suitable for a number of operations including simplification, color texturing, smooth-surface fitting, and rapid prototyping. The meshes generated by the merging process are typically dense, with vertices spaced less than 0.5 mm apart. Objects containing large flat areas may be represented with much fewer triangles, thus the need for mesh simplification algorithms. Color data may also be collected for scanned objects, and one research goal is to derive illumination independent body color and "shininess" for each point on the surface. In addition, the merged meshes are triangular tessellations, but CAD packages for manufacture and animation typically require NURBS surfaces. To address this need, Krishnamurthy & Levoy [1996] have developed a method for interactively creating tensor product B-spline surfaces from dense polygonal meshes. Finally, the models generated by the volumetric method of merging meshes are generally manufacturable with rapid prototyping technology. Near the end of this thesis, we describe the manufacture of one of our models using stereolithography.

1.5 Contributions

The contributions of this thesis are two-fold. First, we describe and demonstrate a new method for optical triangulation, known as spacetime analysis. We show that this new method yields:

- Immunity to reflectance variations
- Immunity to shape variations

We also show that the influence of laser speckle remains a fundamental limit to the accuracy of both traditional and spacetime triangulation methods.

Second, we describe a new, volumetric algorithm for building complex models from range images. We demonstrate the following:

- The method is incremental, order independent, represents sensor uncertainty, and behaves robustly.
- The method is optimal under a set of reasonable assumptions.
- Extending the method to qualify the emptiness of space around an object permits us to construct hole-free models.
- Through careful choice of data structures and resampling methods, the method can be made time and space efficient.

Using the volumetric algorithm to combine the range images generated with our optical triangulation method, we have constructed the most complex computer models published to date. In addition, because these models are hole-free, we are able demonstrate their manufacturability using a layered manufacturing method called stereolithography.

1.6 Organization

Chapter 2 begins with a more detailed description of optical triangulation methods. We then characterize the primary sources of error in traditional optical triangulation and conclude with a discussion of some previous work.

Next, in Chapter 3, we develop a new method for optical triangulation called "spacetime analysis". We show how this method corrects for errors due to changes in reflectance and shape, but is still limited in accuracy by laser speckle.

In Chapter 4, we describe an efficient implementation of the spacetime method, and we demonstrate results using an existing optical triangulation scanner modified to allow digitization of the imaged light reflections. Portions of Chapters 2-4 are described in [Curless & Levoy 1995].

The next four chapters concern the problem of surface reconstruction from range images. In Chapter 5, we describe some previous work in the area, and then provide a mathematical framework that motivates a new method for surface reconstruction. This new method shows that the maximum likelihood surface is an isosurface of a function in 3-space.

In Chapter 6, we develop a new volumetric algorithm for surface reconstruction from range images. This algorithm utilizes the mathematical framework described in the previous chapter and extends it to allow description of the occupancy of space around the surface. This extension leads to a method for generating models without boundaries. In addition, we discuss some of the limitations of the volumetric method and suggest some possible solutions.

The potential storage and computational costs of the volumetric method require implementation of an efficient algorithm. In Chapter 7, we describe one such algorithm and then analyze its asymptotic complexity.

Next, in Chapter 8, we demonstrate the results of reconstructing range surfaces using the volumetric method and the data acquired with the scanning system described in Chapter 4. Portions of Chapters 5-8 are also described in [Curless & Levoy 1996].

Finally, in Chapter 9, we summarize the contributions of this thesis and describe some areas of future work.