View Planning for Automated Three-Dimensional Object Reconstruction and Inspection

WILLIAM R. SCOTT and GERHARD ROTH

National Research Council of Canada

AND

JEAN-FRANÇOIS RIVEST

University of Ottawa

Laser scanning range sensors are widely used for high-precision, high-density three-dimensional (3D) reconstruction and inspection of the surface of physical objects. The process typically involves planning a set of views, physically altering the relative object-sensor pose, taking scans, registering the acquired geometric data in a common coordinate frame of reference, and finally integrating range images into a nonredundant model. Efficiencies could be achieved by automating or semiautomating this process. While challenges remain, there are adequate solutions to semiautomate the scan-register-integrate tasks. On the other hand, view planning remains an open problem—that is, the task of finding a suitably small set of sensor poses and configurations for specified reconstruction or inspection goals. This paper surveys and compares view planning techniques for automated 3D object reconstruction and inspection by means of active, triangulation-based range sensors.

Categories and Subject Descriptors: I.2.9 [**Artificial Intelligence**]: Robotics—*sensors*; I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding—*Modeling and recovery of physical attributes*; I.4.1 [**Image Processing and Computer Vision**]: Digitization and Image Capture—*Scanning*; I.5.4 [**Pattern Recognition**]: Applications—*Computer vision*

General Terms: Algorithms, Design, Measurement, Performance

Additional Key Words and Phrases: View planning, range images, object reconstruction, object inspection

1. INTRODUCTION

The demand for high—quality three dimensional (3D) virtual models of complex physical objects is growing in a wide range of applications (e.g., industrial, training, medical, entertainment, cultural, architectural). Computer graphics can produce synthetic models for some of these

---

This work was partially sponsored by scholarships from the National Science and Engineering Research Council (Canada), the Ontario Graduate Student Scholarship program, and the University of Ottawa graduate scholarship program. The research was conducted while the principal author was a guest researcher at the National Research Council of Canada's Institute for Information Technology.

Authors' addresses: W. R. Scott and G. Roth, Computational Video Group, Institute for Information Technology, National Research Council, M-50, 1200 Montreal Rd., Ottawa, Ont., Canada, K1A 0R6; email: {william.scott, gerhard.roth@nrc-cnrc.gc.ca}; J.-F. Rivest, Defence Research Establishment Ottawa, 3701 Carling Ave., Ottawa, Ont., Canada, K1A 0Z4.

applications by a combination of artistic and technical design processes. In the computer vision field, on the other hand, model acquisition is achieved by measuring the shape and visual properties of real physical objects. Numerous applications require the computer vision approach to object surface reconstruction.

Automating or semiautomating the current manual, time-consuming reconstruction process would improve productivity and quality, reduce operator skill level and rework requirements, and lower cost. While a limited number of commercial systems (Hymarc Ltd., Rapidscan, www.hymarc.com; Digibotics Ltd., Adaptive Scan, www.digibotics.com) offer semiautomated tools to aid scanning, there is presently no general purpose commercial system for automated object reconstruction.

Both mechanical and optical geometric measurement sensors are used for 3D object reconstruction and inspection. Mechanical Coordinate Measuring Machines (CMMs) use precise mechanical movement stages and touch probes to achieve very high measurement precision—around 1 μm for a suitably calibrated high-end machine operated by a skilled technician. However, the acquisition and operating costs for such precision are high. With a relatively low data acquisition rate, CMMs are generally suitable for sparse sampling applications only. While their measurement precision is inferior to CMMs, optical sensors have lower capital and operating costs and are capable of dense, noncontact sampling at high throughput rates.

This survey deals with view planning for automated high-quality object reconstruction or inspection by means of active, triangulation-based range cameras. In this context, view planning is the process of determining a suitable set of viewpoints and associated imaging parameters for a specified object reconstruction or inspection task with a range camera and positioning system. By object reconstruction we mean acquisition of a virtual computer model of the surface of a physical object. This is normally a triangulated polygonal mesh representation of surface geometry. Alternative surface representations such as splines, swept cylinders, and super quadratics are advantageous in special cases but are neither as general purpose nor as flexible as a triangular mesh.

While dealing mainly with object reconstruction, the techniques examined here also apply to inspection. The later is an easier problem, having the advantage of a preexisting model. Essential information can be collected off-line in a suitable data structure, allowing different tasks to be planned quickly online.

Many applications are concerned with shape only. A growing number also seek coregistered surface geometry and surface visual texture or reflectance properties. Capturing high-quality surface geometry (the focus of this survey) is a prerequisite for the accurate measurement of reflectance properties which may be achieved simultaneously with some range sensors [Baribeau et al. 1991].

The paper is structured as follows. After an overview of the problem (Section 2), view planning requirements are defined in detail (Section 3). We then reference related surveys in the field (Section 4) before presenting a survey of view planning methods in two broad categories: model-based (Section 5) and non-model-based (Section 6). Finally, we compare (Section 7) and critique (Section 8) existing methods with respect to the defined requirements, examine related issues (Section 9), and conclude with a discussion of the remaining open problems (Section 10).

2. PROBLEM OVERVIEW

2.1. Imaging Environment

The imaging environment (Figure 1) for object reconstruction consists of a range camera, positioning system, various fixtures, and the target object.

Range camera. The principal component of the imaging environment is a range camera—a sensor for 3D shape measurement. A wide variety of technologies are used for measuring object shape [Besl 1989]. Range cameras can be categorized

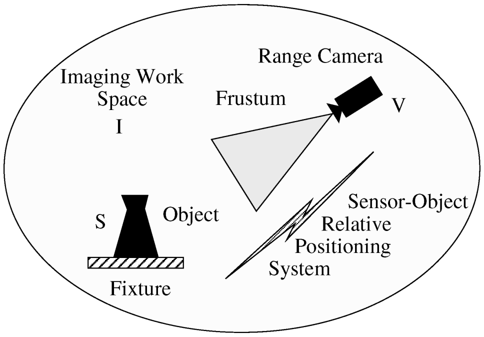

Fig. 1. Imaging environment.

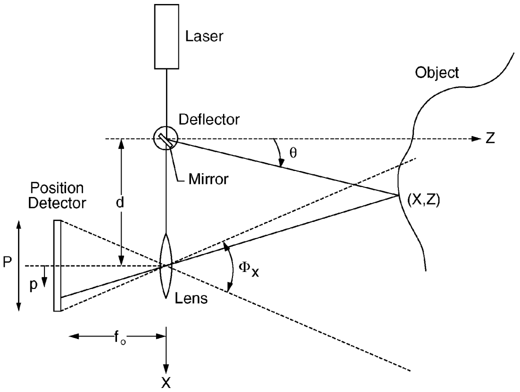

Fig. 2. Conventional active triangulation. (From El-Hakim and Beraldin[1994]; ©IEEE 1994.)

into active and passive devices. The most common passive range-finding technology, stereo vision, can provide accurate measurement at short stand-off distances. However, passive stereo depends on visual texture on the target object, is subject to illumination constraints and interference, and can provide only sparse depth measurement. With an integral illumination source, active sensors are capable of dense range measurement and are less susceptible to external interference. However, the illumination source (frequently a laser) may impose system safety constraints. Active range sensors can be divided into two subcategories, time-of-flight and triangulation. Time-of-flight systems require very accurate timing sources and typically provide modest resolution (generally centimeter but occasionally millimeter accuracy) for longer range applications, that is, tens to thousands of meters [Amann et al. 2001]. They are best suited to distance measurement and environment modeling at medium to long ranges. Many different types of triangulation sensors have been developed and are in wide usage. All are based on the principle of triangulating a measurement spot on the object from a physically separate camera optical source and detector. By simple geometry, the x, z coordinates of the illuminated spot on the object are calculated (Figure 2). In general, active triangulation-based range cameras are capable of very precise ($\leq$100 μm), dense depth measurements (many samples per square millimeter) over relatively small sensor frustums up to about a meter in standoff distance [El-Hakim and Beraldin 1994]. In the following, z is range, x is distance across the scan, f_0 is the sensor focal length, p is the position of the imaged spot on the sensor detector, and θ is the laser scan angle:

$$z = \frac{d f_0}{p + f_0 \tan\theta}, \tag{1}$$

$$x = z \tan\theta. \tag{2}$$

This survey focuses on triangulation-based active laser range scanners as they are widely used for precise measurement for both object reconstruction and inspection. Several sensor attributes impact view planning. The optical baseline, the distance between the laser and the optical receiver, is significant with respect to the measurement stand-off distance. Consequently, coverage shadow zones arise with respect to visibility by the laser, the optical receiver, or both. The sensor field of view, depth of field, and therefore frustum volume are limited. Measurement precision and sampling density are highly nonuniform within the frustum. Additionally, measurement is subject to random nonisotropic geometric noise [Beraldin et al. 1993] and several artifact phenomena [Curless and Levoy 1996].

Positioning system. Imaging all sides of an object requires a variety of viewing perspectives. Thus, a positioning system is required to move the sensor, the object, or both in some combination. An imperfect positioning system introduces pose error, which has two consequences. First, a planned view is not the one actually executed. Sufficiently severe pose error can render view planning futile. Second, the resulting pose uncertainty necessitates an image registration stage.

Viewpoint space. Viewpoints should be treated as generalized viewpoints $(\mathbf{v}, \lambda_\mathbf{s})$ consisting of sensor pose $\mathbf{v}$ and a set of controllable sensor parameters $\lambda_\mathbf{s}$ [Tarabanis et al. 1995b; Roberts and Marshall 1997; Pito 1997a]. Thus each viewpoint has an associated sensor configuration. Sensor pose is limited by the degrees of freedom and range of motion of the positioning system. Viewpoint space V is the set of generalized viewpoints defined by the range and sampling of these parameters.

Object. The object is the feature to be modeled. Specifically, we wish to capture its 3D shape. Object surface space S is the set of 3D surface points sampled on the object. These can be considered as vertices in the resulting object mesh model.

Fixtures. The object must be supported by a fixture of some kind. Consequently, acquisition of all-aspect views will require the object to be repositioned on the fixture at least once and more likely several times. After each repositioning, the new relative pose must be determined to bring the system into a single coordinate frame. Fixtures, the positioning system, and any other structures in the imaging work space I introduce occlusion and collision avoidance considerations.

2.2. Reconstruction Cycle

The classical model building cycle (Figure 3) consists of four main phases—plan, scan, register, and integrate. A sequence of views, the *view plan* or *next-best-view (NBV) list* N, must be computed. The sensor must be moved to the appropriate pose and configured with appropriate settings, after which scans are taken. Unless the positioning system is error-free, acquired range images must be registered by an image-based registration technique such as the standard Iterative Closest Point (ICP) method [Besl and McKay 1992]. Subsequently, registered images must be combined into a single, nonredundant model—a process commonly known as *integration*. The process continues until some stopping criteria are satisfied.

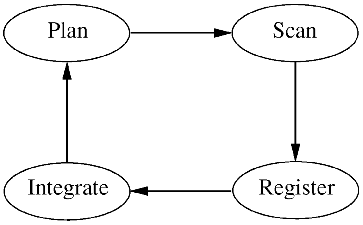

Fig. 3. Object reconstruction cycle.

These are the basic reconstruction steps. Other related activities are briefly as follows. Calibration of both sensor and positioning system is an essential prerequisite and may need to be repeated after certain positioning system reconfigurations. Filtering of noise and artifacts is required at several stages. Depending on the application, there may also be a requirement for model compression or decimation and for texture mapping of reflectance data onto the geometry [Baribeau et al. 1991].

While challenges remain, adequate solutions exist to automate the scan-register-integrate functions.[1] However, automated view planning remains an open problem despite two decades of research. For large reconstruction tasks, the image registration and integration (model building) functions can be very time consuming. By the time deficiencies in the partial model are discovered, the imaging team may have left the site or access to the object may no longer be readily available. Hence, there is a need for a view planning

---

[1] Research on ICP refinements remains an active field of research. For example, see Rusinkiewicz and Levoy [2001] and Langis et al. [2001]. Research on model building also remains active, with a number of open issues [Roth 2000].

scheme capable of developing reliable view plans in a timely manner for comprehensive object coverage in accordance with all technical criteria of the reconstruction task.

A different approach is to automate only the register-integrate functions and rely on human-based view planning and scanning. Huber [2001] used graph-based methods to increase the reliability of multiview registration and integration. Hébert [2001] used a hand-held scanner to sweep the object as one would use a paint sprayer. Both techniques employ progressive model building. Both require near real-time registration and integration, difficult tasks whose computational complexity is sensitive to model size and number of images taken. Registration of single 1D scans faces reliability and stability issues. Humans are relatively good at high-level view planning for coverage of simple objects but even experienced operators will encounter considerable difficulty with topologically and geometrically complex shapes. A visualization feedback loop can overcome some of these difficulties. Hand-held or even robotically supported sensors will introduce pose error with consequences discussed later in this review. And, of course, heavy reliance on a human operator defeats the ultimate objective of full automation. Notwithstanding these weaknesses, human-based view planning techniques are good candidates as interim semiautomated measures for applications with modest quality objectives. The balance of this review is restricted to machine-based view planning.

2.3. The View Planning Problem

From the foregoing, it is apparent that the view planning problem (VPP) involves reasoning about the state of knowledge of three spaces—object surface space S, viewpoint space V, and imaging work space I. While tied to 3D surface geometry, S is amenable to 2D parameterization. In the unrestricted case, the pose component of viewpoint space V is six dimensional—three position and three rotation. Configurable sensor parameters such as laser power and scan length can further raise the dimensionality of generalized viewpoint space, which we indicate as $6D^+$. Imaging workspace I is the 3D region capable of being viewed by the sensor over the full range of the positioning system. It is of concern primarily for visibility analysis and collision avoidance considerations. The complexity of the VPP is immediately apparent from the high dimensionality of the search spaces in S, V, and I.

Given this imaging environment and reconstruction cycle, the view planning task is deceptively simple, yet computationally complex. Expressed informally, the view planning problem is as follows:

> For a given imaging environment and target object, find a suitably short view plan N satisfying the specified reconstruction goals and achieve this within an acceptable computation time.

N will be a subset (preferably a very small subset) of viewpoint space V, that is, $N \subset V$. As the View Planning Problem has been shown to be NP-Complete ([Tarbox and Gottschlich 1995]; [Scott 2002]), in most realistic reconstruction or inspection tasks it is impractical to seek an absolute minimal length view plan. What is an "acceptably short" computation time is application dependent. In an industrial setting, object reconstruction involves repetitive modeling of different objects. In this case, throughput is critical, emphasizing time-efficient algorithms. Perhaps some aspects of quality, such as the view plan length, may be traded off in the interests of computational efficiency. In some reconstruction tasks, the primary objective is model quality and scanning efficiency is secondary. Inspection applications involving repetitive execution of an inspection plan on a production line place a premium on view plan efficiency over the time taken to create it.

Several other machine vision applications also involve view planning, such as environment modeling [Sequeira and Gonçalves 2002], autonomous exploration [Tremblay and Ferrie 2000], and sensor-based robot path planning [Yu and Gupta 2000]. View planning requirements, techniques, and sensors appropriate for these applications are somewhat different than for high-quality object reconstruction/inspection. These applications typically have much lower accuracy and sampling density requirements but potentially larger models, so different sensors may be appropriate such as time-of-flight (optical, sonar or radar) as well as passive optical technology.

2.4. Performance-Oriented Reconstruction

High-quality model building requires view planning for *performance-oriented* reconstruction [Scott et al. 2000], which is defined as model acquisition based on a set of explicit quality requirements expressed in a *model specification*. High-quality inspection tasks are also generally specification-driven. In addition to all-aspect coverage, measurement quality may be specified in terms of precision, sampling density, and perhaps other quality factors. Specified measurement quality may be fixed or variable over the object surface.

Performance-oriented view planning requires suitable models of both sensor and positioning system performance. These can be combined in an imaging environment specification. Specifically, it requires the following:

—a *sensor model* with a description of the camera geometry and frustum geometry and a characterization of measurement performance within the calibrated region; and

—a *positioning system model* describing the degrees of freedom, range of motion, and positioning performance within the movement envelope.

There has been relatively little work on performance-oriented view planning. Cowan and Kovesi [1988] used a constraint satisfaction approach for sensor location subject to task requirements which included resolution, focus, field of view, visibility, view angle, and prohibited regions. The MVP system developed by Tarabanis et al. [1995] automatically synthesizes views for robotic vision tasks with intensity cameras based on a task specification as well as a geometric model of the scene and optical models of the sensor and illumination sources. The MVP task specification includes visibility, field of view, resolution, focus, and image contrast. Soucy et al. [1998] described a system for automatically digitizing the surface of an object to a prescribed sampling density using a contour following scheme. Prieto et al. [1999, 2001] set CAD-based inspection criteria based on range sensor performance characterization. A number of authors have considered grazing angle as a subjective quality measure. Otherwise, the objective of most work in the field has been explicitly or implicitly limited to full surface coverage. Recently, Scott et al. [2000, 2001b] used a model-based, multistage approach to performance-oriented view planning based on an input specification for sampling precision and density.

3. VIEW PLANNING REQUIREMENTS

In this section, we define requirements and performance measures for the view planning process in greater detail.[2] The requirements are drawn up from a research perspective. Additional development and system integration details will arise when moving prototype view planning algorithms to a production configuration of an automated object reconstruction or inspection system.

3.1. Assumptions

To begin, our definition of the view planning problem for object reconstruction and inspection is based on several assumptions:

—Both the range camera and positioning system are calibrated and calibration parameters are available to the view planning module.

—The imaging work space is compatible with object size and shape.

---

[2] Pito [1997a] also provided a good description of the view planning problem and defined constraints on the choice of the next-best-view and desirable properties of view planning algorithms.

—Range camera performance is compatible with the model specification for the reconstruction task.

—The sensor and positioning system have compatible performance. Specifically, pose error is compatible with the volume of the range camera frustum.

—Object shape is compatible with sensor capabilities.

—Object material properties are compatible with sensor capabilities.

The last two items merit elaboration. Range cameras are inherently sampling devices. Therefore, the reconstruction process is subject to the limitations of a sampled representation of continuous surface shape. The familiar sampling theorem for 1D and 2D signals does not extend to three dimensions. As there is no natural low-pass filter for 3D physical shape, aliasing effects can be anticipated. Convoluted shapes can easily be conceived which cannot be fully imaged by a range camera, for example, a gourd or a sea urchin-like shape. Deep cavities, deep holes, or multiple protrusions will be troublesome or completely impractical to image. Therefore, practicality requires that object geometry and topology be "reasonable."

Laser scanners have difficulty with certain materials. They perform best with material having surface scattering properties and a relatively uniform reflectance, neither overly absorbent nor overly reflective. Natural or humanly made material with volume scattering properties such as hair and glass produce multiple reflections over a surface depth zone, degrading or prohibiting accurate range measurement.

The dynamic range of the current generation of range sensors is limited. Excessively absorbent material scatters insufficient light, causing "drop outs," that is, missing data. Shiny surfaces are characterized by specular reflection. Depending on illumination and observation geometry, specular surfaces may result in drop outs due to insufficient scattered energy or outliers or drop outs (depending on camera design) due to receiver saturation from excessive received energy. Multiple reflections on a shiny surface in corner regions can also produce wild measurements. Finally, abrupt reflectance changes will produce edge effects in the form of measurement biases and artifacts [Curless and Levoy 1996].

Table I. View Planning Requirements

Category	Requirement
General	Model quality specification
	Generalizable algorithm
	Generalized viewpoints
	View overlap
	Robust
	Efficient
	Self-terminating
Object	Minimal a priori knowledge
	Shape constraints
	Material constraints
Sensor	Frustum
	Shadow effect
	Measurement performance
Positioning System	6D pose
	Pose constraints
	Positioning performance

Consequently, the current state-of-the-art is limited to view planning for objects with "reasonable" shapes and benign reflectance characteristics. Some view planning techniques are applicable to restricted shape classes, such as manufactured objects constructed of a limited set of shape primitives. The issue of extending view planning to objects with more difficult reflectance characteristics is addressed in Section 10.1.

3.2. Constraints and Requirements

The following constraints and requirements apply to the view planning task for object reconstruction. They are summarized in Table I under the categories of general, object, sensor, or positioning system. The following amounts to a "research specification" for a view planning system. Where objectives are quantified, it is presumed that research-quality algorithms are subject to refinement, development, and system integration on suitable production hardware and software platforms.

3.2.1. Model Quality Specification. By definition, performance-oriented view planning is based on a model specification containing explicit, quantified, model quality requirements such as measurement precision and sampling density. For object reconstruction, there is usually an implicit requirement for 100% surface coverage. In other applications such as inspection, it may be appropriate to limit coverage to specified object regions.

3.2.2. Generalizable Algorithm. For wide applicability, the technique should apply to a broad class of range sensors, positioning systems, and objects.

3.2.3. Generalized Viewpoints. In addition to sensor pose, the view planning algorithm should plan reconfigurable sensor parameters such as laser power and scan length through the use of generalized viewpoints.

3.2.4. View Overlap. The algorithm should provide the degree of image overlap necessary for integration and registration. Integration requires image overlap along the boundaries of adjacent images in the view plan. If pose error is such that image-based registration is required, the image set should have sufficient shape complexity in overlapping regions for registration of the image set within the specified precision.

3.2.5. Robust. The view planning algorithm should be immune to catastrophic failure, handle sensor and positioning system noise and artifacts, and require minimal operator intervention.

3.2.6. Efficient. The view planning algorithm should be sufficiently efficient to be competitive with skilled human operators. As an initial goal, the algorithm should be capable, with production-quality hardware and software, of producing a specification-compliant view plan for a moderately complex object within 1 h.

3.2.7. Self-Terminating. The view planning algorithm should recognize when the goal has been achieved or when progress toward it has stalled.

3.2.8. Limited a priori Knowledge. The view planning algorithm should be effective with minimal a priori object knowledge, specifically no more than approximate bounding dimensions and centroid.

3.2.9. Shape Constraints. The view planning algorithm should be effective with all reasonable object geometry and topology that is compatible with sensor measurement capabilities.

3.2.10. Material Constraints. The algorithm should be effective with all object material properties compatible with sensor measurement capabilities.

3.2.11. Frustum. Sensor frustum shape should be modeled. This includes sensor depth of field, field of view, and scan length or scan arc.

3.2.12. Shadow Effect. The bistatic nature of the sensor should be modeled—that is, the physical separation of the laser and detector.

3.2.13. Measurement Performance. Sensor measurement performance should be modeled, including variation of measurement precision and sampling density within the frustum and surface inclination effects. The following range camera artifacts should be modeled: geometric and reflectance step edges and multiple reflections.

3.2.14. 6D Pose. The view planning algorithm should handle a positioning system with an unconstrained pose space-three position and three rotation.

3.2.15. Pose Constraints. The view planning algorithm should model constraints on the degrees of freedom and range of motion of the positioning system.

3.2.16. Positioning System Performance. Positioning system performance should be modeled to include pose error and repositioning/reconfiguration time.

3.3. Performance Measures

To date, adequate standards for quantifying view planning performance have not existed. This complicates the comparative assessment of view planning algorithms. We propose the following measures for evaluating view planning algorithm performance for a given reconstruction task with a given imaging environment:

- *View plan quality* The quality of a view plan is determined by the quality of the reconstruction it generates. This can be expressed as the overall verified measurability[3] m_v of the reconstruction with respect to the model specification. m_v is an indicator of the fidelity and robustness of the view planning process plus the adequacy of the discretization schemes for viewpoint space and surface space. The goal is $m_v = 1.0$.
- *View plan efficiency* A measure of view plan efficiency is the length of the generated view plan relative to the optimum achievable for that task, that is, $e_v = n_{Opt}/n$, where $n = |N|$. As it may be impractical to determine with certainty the optimum view plan length n_{Opt} for complex reconstruction tasks, a suitable surrogate is the length of the best solution n_{Best} found thus far among all view planning techniques examined for the same task, that is, $n_{Opt} \approx n_{Best}$. e_v is an indicator of the efficiency and completeness of discretization schemes for viewpoint and surface space and the efficiency of the set covering process. The goal is $e_v = 1.0$.
- *View plan computational efficiency* Measures of the computational cost of generating the view plan are computational complexity and execution time on a defined platform. While each has weaknesses, together they provide a reasonable appreciation of algorithm performance. Computational efficiency indicates the efficiency of discretization schemes for viewpoint and surface space, visibility analysis, and the set covering process. Computational efficiency goals are application-dependent.

4. RELATED SURVEYS

Surveys relevant to view planning for object reconstruction can be found primarily in doctorate and masters theses published over the last decade (Table II).

In a brief and clearly written technical report addressing only seven papers (Banta et al. [1995], Connolly [1985], Hutchinson and Kak [1989], Maver and Bajcsy [1993], Pito [1996b], Tarabanis et al. [1995a], and Whaite and Ferrie [1990]), Hall-Holt [1998] quickly zeroed in on two key issues—representation of available geometric information and how to deal with the potentially vast search space. While he did not answer the questions he posed, Hall-Holt succinctly examined from several perspectives the view planning approaches taken by some of the most notable authors. The survey concluded with a discussion of open issues: algorithm efficiency in the presence of large amounts of data, appropriate geometrical representations, hybrid models, multiresolution techniques, ray tracing avoidance, quantitative performance assessment, performance benchmarks, and improved scanner models.

The classic survey by Tarabanis et al. [1995a], considered three vision tasks—inspection, recognition, and reconstruction—but focused almost exclusively on the first of these. The underlying perspective was that of conventional intensity imaging, although a few references were made to range imaging. View planning was categorized as following one of three paradigms—synthesis, generate

---

[3] Verified measurability of a reconstruction can be seen as the ratio of reconstructed surface area compliant with the specified model criteria to the overall object surface area. For more precise definitions, refer to Scott et al. [2001b, 2002].

Table II. View Planning Surveys

Year	Author(s)
2002	William Scott [2002] Ph.D. thesis
1999	Flavio Prieto [1999] Ph.D. thesis
1998	Michael Reed [1998] Ph.D. thesis
1998	Olaf Hall-Holt [1998] Technical Report
1997	Steven Abrams [1997] Ph.D. thesis
1997	Brian Curless [1997] Ph.D. thesis
1997	Nikos Massios [1997] Masters thesis
1997	Dimitri Papadopoulos-Orfanos [1997] Ph.D. thesis
1997	Richard Pito [1997a] Ph.D. thesis
1997	Roberts and Marshall [1997] technical report
1997	Yiming Ye [1997] Ph.D. thesis
1996	Joseph Banta [1996] Masters thesis
1996	Leland Best [1996] Ph.D. thesis
1996	Vítor Sequeira [1996] Ph.D. thesis
1996	Mark Wheeler [1996] Ph.D. thesis
1995	Jasna Maver [1995] Ph.D. thesis
1995	Tarabanis et al. [1995] survey paper
1994	Stephane Aubry [1994] Ph.D. thesis
1994	Kiriakos Kutulakos [1994] Ph.D. thesis
1994	Chris Pudney [1994] Ph.D. thesis
1994	Lambert Wixson [1994] Ph.D. thesis
1993	Sergio Sedas-Gersey [1993] Ph.D. thesis
1993	Glen Tarbox [1993] Ph.D. thesis
1993	Xiaobu Yuan [1993] Ph.D. thesis
1992	Konstantinos Tarabanis [1992] Ph.D. thesis
1992	Besma Roui-Abidi [1992] Ph.D. thesis
1990	Michael Buzinski [1990] Masters thesis
1990	Seungku Yi [1990] Ph.D. thesis

and test, or expert system. The open problems in view planning circa 1995 as seen by Tarabanis et al. [1995a] were the following:

—modeling and incorporating other constraints such as integrating collision avoidance with view planning, dynamic sensor planning, and modeling the operating range of the employed sensor;

—modeling and incorporating other sensors such as tactile, range, force-torque, and acoustic sensors;

—relaxing some of the assumptions made in current approaches such as feature uncertainty and error characterization;

—illumination planning to include higher-order lighting and reflectance models, multiple sources, specularity, and interreflections; and

—sensor and illumination modeling to include mapping model parameters to controllable sensor settings and accurate sensor noise models.

A comprehensive survey of automated visual inspection systems by Newman and Jain [1995] considered binary, intensity, colored and range image sensors for a wide range of inspection applications.

These surveys categorized view planning methods in several different ways.

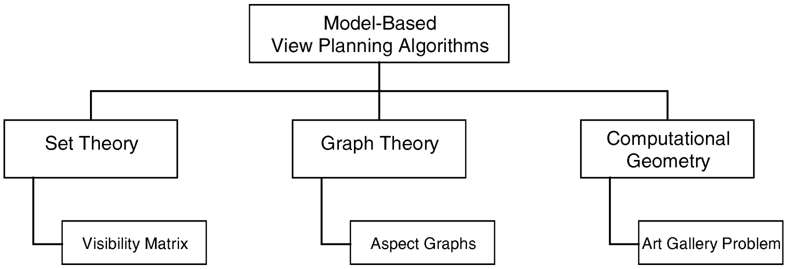

Fig. 4. Model-based view planning algorithms.

From our perspective, we begin with two top level categories—model-based or non-model-based. It is convenient to then sub-categorize the two classes somewhat differently.

5. MODEL-BASED VIEW PLANNING

Model-based view planning methods can be categorized by the representation used for the knowledge embedded in the object model (Figure 4). "Model-based" methods base planning on an a priori object model at some level of fidelity.

5.1. Set Theory Methods

Visibility matrices. At the heart of the set theoretic approach is a visibility matrix, whose elements encode in a single data structure the visibility of discrete object surface points from each pose in quantized viewpoint space. A more sophisticated enhancement replaces visibility by measurability, in which the data elements encode estimated measurement quality.

Tarbox and Gottschlich. Tarbox and Gottschlich [1995] incorporated elements of set theory, graph theory, computational geometry, mathematical morphology, and statistical nonlinear optimization in their thorough examination of view planning for automated inspection. They discretized viewpoint space by recursive subdivision of an icosahedron and constrained solutions to lie on a viewing sphere completely containing the object with the viewing direction oriented toward the center of the sphere. They further assumed that the object lies completely within the sensor's frustum. An octree-encoded voxel occupancy model represented the sensing volume. Their view planning approach was tailored to a specific long-baseline, active triangulation-based range sensor so that the view planning problem becomes one of examining the set of all ordered pairs of points on the view sphere separated by the specified sensor baseline. They briefly discussed culling inadmissible points from the viewpoint set by ray tracing operations, such as those occluded by a mounting fixture and supporting plane. The principal consideration in viewpoint selection is taken to be the surface area that a given viewpoint is capable of sensing.

Consideration of grazing angle effects figured prominently in their approach. Different grazing angle thresholds were set for the light source and the camera. The only grazing angle effects considered were the impact on dilution of surface sampling density and viewability at acute angles in the presence of surface microstructure. This masked the most important impact of grazing angle which is on measurement error. Other aspects of sensor noise and artifacts were not considered.

The authors noted that it would be desirable to find the shortest possible view plan. However, as this problem is known to be NP-complete, the authors concluded that it would be necessary to employ a view planning algorithm that can find a satisfactory but not necessarily optimal view sequence.

Tarbox and Gottschlich [1995] developed and examined the performance of three algorithms. The first two employed an irrevocable selection strategy and differed by the manner in which grazing angle constraints were treated. The third algorithm was novel in that it employed a revocable selection strategy for viewpoint set minimization based on a randomized search with simulated annealing.

To account for slight pose errors, they removed viewpoints which were not robust to pose variation. This was accomplished by morphological operations on the viewpoint set associated with each surface point.

Their approach was based on a measurability matrix $\mathcal{M}(i, j)$ computed by a complete visibility analysis for the laser source and camera over the set of all surface points and all admissible viewpoints. Given the span of the discretized variables, the computational complexity of the approach was prohibitive. This was a fundamental limitation regarding direct applicability of the work to object reconstruction. Nevertheless, the authors' thorough and well-written treatment of the subject provided useful insights into several aspects of the view planning problem.

5.2. Graph Theory Methods

Aspect graphs. An aspect graph of an object has a node representing every aspect of that object and arcs connecting all adjacent aspects on an object [Tarbox and Gottschlich 1995; Bowyer and Dyer 1990]. An aspect is loosely defined as the set of viewpoints of the object such that the unoccluded portion of the object seen from all those viewpoints is qualitatively the same. In other words, viewpoint space is partitioned into regions providing equivalent views. Each node represents one such region while arcs represent their adjacency in viewpoint space.

While aspect graphs are an intriguing theoretical possibility, there are practical difficulties. When dealing with objects as simple polyhedra and considering a viewpoint purely in terms of the visibility of object features from that point, the criterion "qualitatively the same" simply means that the same vertices, edges, and faces are visible from that viewpoint. When the object is changed from discrete to continuous, the quantization of viewpoint space inherent to an aspect graph representation strictly fails. For some applications, one may redefine an aspect graph in terms of the visible topology of the object's silhouette, but this is not useful for object reconstruction. Additionally, as we are dealing with a sensing problem which involves not only visibility but mensuration, viewpoints with the same visibility (i.e., the same aspect) are by no means equivalent for sensing purposes. Second, aspect graphs for even moderately complex objects quickly become huge and unwieldy, as does the computational complexity of handling them. Consequently, aspect graph theoretical development is not sufficiently mature [Faugeras et al. 1992] to be a suitable representation basis for performance-oriented view planning.

5.3. Computational Geometry Methods

Art gallery problem. A classic 2D computational geometry problem concerns the visibility of edges in a polygon, the so-called "art gallery problem" [Urrutia 2000; Xie et al. 1986]. Given the floor plan of an art gallery as a polygon P, the problem is to find an upper bound on the number of "guards" represented by points such that the interior walls of P are completely visible [Kahn et al. 1980]. The task is to determine the minimum number and placement of guards who collectively can see all walls in the gallery. The object reconstruction problem is somewhat related to this classic computational geometry problem. However, there are additional complexities—three dimensions, bistatic visibility (source and receiver), plus mensuration in lieu of visibility.

6. NON-MODEL-BASED VIEW PLANNING

It is convenient to classify existing view planning methods (most of which are non-model-based) by the domain of reasoning about viewpoints—that is, *volumetric*, *surface-based*, or *global* (Figure 5). Some methods combine several techniques. The majority fall into two subcategories—voxel occupancy or occlusion edges.

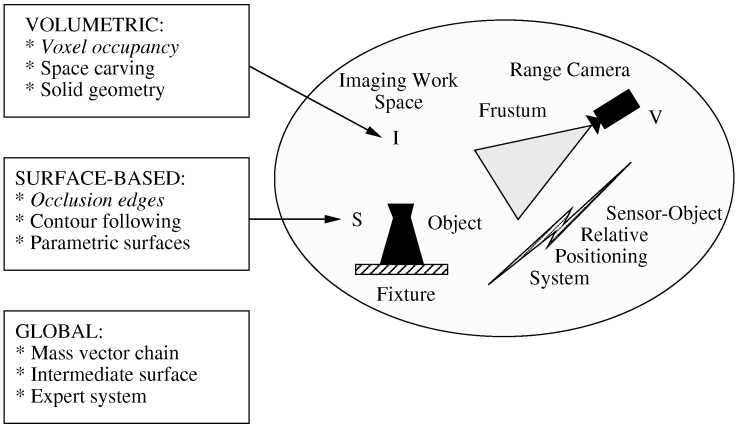

Fig. 5. Traditional non-model-based view planning methods.

6.1. Surface-Based Methods

Surface-based algorithms reason about knowledge of object surface space S.

6.1.1. Occlusion Edge Methods

Occlusion edges. The most commonly used traditional method exploits geometric jump edges. Illustrated in Figure 6, this approach is based on the premise that occlusion edges internal to the image indicate surface areas not yet sampled, while boundary jump edges represent the boundary of the unobserved volume. In both cases, occlusion edges provide cues for the next-best-viewing direction.

Maver and Bajcsy. Some of the earliest papers using the occlusion edge method were by Maver and Bajcsy [1990, 1993]. Their approach was tailored to a long baseline laser profile scanner with a positioning system limited to a single rotational degree of freedom. The two-stage method separately considered source and receiver occlusions. The focus was on finding and labeling boundary occlusion edges. View planning was done with reference to a plane defined by the support fixture. The initial stage planned views as a rotation about the range camera boresight. The projection of occluded regions on the support plane was approximated by polygons. Occlusion-free arcs were computed by a visibility analysis making assumptions about the height of each “shadow zone pixel.” A histogram was formed by summing the viewing arcs for each pixel and a next-best-view determined from histogram maxima. The algorithm was computationally intensive due to the many-on-many visibility calculations. Performance was sensitive to the underlying assumptions.

In a second view planning stage, described analytically but not implemented, occlusion edges of source shadow zones were used in a similar manner to determine a new orientation for the camera illumination plane. The authors restricted the search for new scanning planes to the set perpendicular to the original scanning set up. Simulated results were shown.

The approach was limited by its formulation as a $2\frac{1}{2}$ D scene view rather than a true 3D perspective. The scene was defined as a horizontal x-y array of pixels with a height value, either measured or assumed. Coupled with sensor and positioning system limitations, the method was incapable of acquiring an all-aspect model. Garcia et al. [1998a, 1998b] also utilized an occlusion edge method.

6.1.2. Contour Following. Once having located a portion of the object, the contour following technique involves “painting” the object with the sensor by keeping it in close proximity to the surface at all times. The technique has been applied to a class of range sensors with a limited sensing volume. Collision avoidance is a primary concern. Pudney [1994] described the application of contour following to a robot-mounted range sensor.

Conventional terminology labels range sensors as range "cameras," implying an output in the form of an "image." In fact, most range cameras are profile scanners which only acquire data in an image format through linear or rotational sensor motion by a fixed or programmable amount. Contour following usually involves acquiring long strips of profiles of an arbitrary length rather than a range image in the conventional sense.[4] Because the positioning systems used typically are accurate in pure translation but have limited mobility or may require recalibration in rotation, scanning is frequently done along nonlinear curves with a fixed orientation. The technique is widely used commercially with mechanical CMM devices. Contour following works best with objects large in size relative to sensor coverage and for relatively simple shapes with smoothly flowing lines. It is more problematic with smaller objects and more complex shapes. Other contour following work with range sensors includes Soucy et al. [1998], Lamb et al. [1999], and Milroy et al. [1996]. The latter used a combined region growing and contour following strategy.

6.1.3. Parametric Surface Representations.

Superquadric models have been widely used in machine vision as flexible, compact representations. However, they are limited to simple scenes unless the scene is segmented and piecewise fitted with separate models. Additionally, superquadrics are highly nonlinear.

Whaite and Ferrie. Related more to robotic exploration than object reconstruction, the autonomous exploration approach of Whaite and Ferrie [1990, 1991, 1992, 1997] nevertheless provides an interesting high-level perspective on characterization and exploitation of uncertainty in view planning. Their gaze-planning strategy used model uncertainty as a basis for selecting viewpoints. After the data is segmented, the task is to find superellipsoid parameters best describing each segmented part. They observed that the best sensor locations are those where the ability to predict is worst—that is, where variance in the fit of the data to the current model is greatest.

Because of concerns about the validity of the linearized theory, they took a conservative approach in which the sensor was always moved in small steps. Using a numerical analysis technique, they searched for and moved the sensor in the direction of maximum uncertainty. Movements were constrained to a view sphere at a fixed geodesic distance from the current scanner location. Experiments showed the algorithm to be attracted to regions of high object curvature.

The authors used a multilayered system design. A general-purpose, sensor-independent, and environment-independent lower-layer navigator module handled view planning. A higher-level explorer module contained application-specific knowledge and handled executive-level functions such as delegating tasks, monitoring progress, making decisions, and resolving conflicts. While the layered design is attractive, it is difficult to see how detailed sensing and environmental factors can be separated from the act of sensor view planning.

The method directly incorporates measurements of sensor noise, although it is unclear exactly which noise sources are modeled or what the fidelity of the corresponding models is. View overlap is a consequence of the conservative nature of the search strategy rather than application of an explicit constraint.

While innovative and providing useful insights into view planning at a high-level of abstraction, the technique was designed more for approximate volumetric modeling of simple scenes for robotic manipulation than for high-definition surface modeling.

---

[4] See also the discussion in Section 9.2.1.

6.2. Volumetric Methods

Volumetric methods select viewpoints by reasoning about the state of knowledge of imaging work space I. Each scan labels a portion of I. The NBV is the viewpoint offering the greatest prospective reduction in uncertainty about I. Volumetric methods focus particularly on the solid shadows cast by the scanned object.

6.2.1. Voxel Occupancy Methods. Common volumetric methods involve encoding space occupancy by a voxel occupancy grid or an octree.

Voxel occupancy grids. Voxelization is a widely used, compact means of encoding spatial occupancy. Voxel grids can also be used for a coarse surface representation, although they are clearly not suited for high-precision modeling. Their principal disadvantage is the large memory requirement for even moderate volumetric quantization levels. In a typical imaging environment, the object occupies only a small portion of the imaging work space and its surface intersects an even smaller portion of the voxelized space. Most authors have ignored the impact of misalignment of spatial quantization intervals [Greespan 2002] between views. The phenomenon is similar to timing jitter in conventional time domain signal processing.

Banta et al. Banta and Abidi [1996] and Banta et al. [1995] defined the NBV as “the next camera pose which will extract the greatest amount of unknown scene information.” Their objective is to minimize the number of views required. The work was similar to that of Connolly [1985] although they constrained the NBV search to a smaller search space. Both the imaging work space and object surface are represented by voxel occupancy grids in which voxels are labeled *occupied* or *unoccupied*. Rather than allowing an “unknown” state, this binary approach labels all points not currently visible as occupied and merges views by a voxel-wise logical “AND” operation. The papers explored several cuing mechanisms for suggesting feasible views, including orienting the sensor in the direction of the viewpoint with the greatest number of potentially visible voxels based on local surface exposure,[5] the viewpoint revealing the greatest number of hidden voxels based on a ray tracing visibility analysis, the mean of the three largest jump edges in the most recently acquired image, or the centroid of the cluster containing the largest number of unknown voxel faces.

The notion of applying several view planning concepts in combination under intelligent control is a useful contribution. The authors also proposed tests to ensure selection of a “good” NBV such as validating by ray tracing that the feature of interest is actually visible and enforcing a minimum angular separation between views on the view sphere. Additionally, they examined several termination criteria, that is, terminating when the size of either the surface model or the occluded model ceases to change by a significant amount or when the ratio of the size of the surface model to that of the occluded model is “large.” However, the proposed termination criteria do not relate to objective performance requirements for the target model.

The various algorithms attempted were effective in what could be called exploratory, approximate view planning for topologically complex objects with purely synthetic data but were less effective in acquiring smaller geometric detail. While innovative, the work was subject to a number of simplifications and limitations. The range camera was treated as an error-free monostatic sensor with uniform performance within the imaging work space. Pose error was ignored. Viewpoints were restricted to a fixed-radius view sphere.

Massios and Fisher. A paper [Massios and Fisher 1998] summarizing Massios’ [1997] Master’s thesis built on previous voxel occupancy methods by using the weighted sum of visibility and “quality”

---

[5] This is similar to Connolly’s “normal” algorithm which fails in concave regions.

factors as an NBV objective function:

$$f_{total}(\vec{v}) = w_v\, f_{visiblity}(\vec{v}) + w_q\, f_{quality}(\vec{v}). \quad (3)$$

The visibility cue was taken from an occlusion edge analysis finding occlusion plane voxels which are defined as unseen voxels with an empty neighbour. Visibility factor $f_{visiblity}$ is set to the number of occlusion plane voxels visible from viewing direction $\vec{v}$, as determined by ray tracing. The absolute value of the dot product of the estimated local surface normal and viewing direction vector is taken as a local quality measure for occupied voxels. A region quality estimate is also formulated for each voxel due to unspecified system inaccuracies. Quality factor $f_{quality}$ is formulated to maximize the number of low-quality voxels visible from a given viewpoint.

Viewpoint space is taken to be a constant radius tesselated sphere, obtained by recursive subdivision of an icosahedron, fitted over the volumetric representation for the object. Using an experimental configuration whose positioning system was limited to a single degree of freedom, the authors showed separate and combined plots of the objective function as a function of angle and the cumulative number of views taken, a presentation format usefully illustrating the impact of the quality measure.

While the introduction of a quality term in the objective function was a valuable contribution, the chosen measure was subjective and did not relate to objective measurement constraints on the reconstructed model. Further, the model was deterministic and considered only one error mechanism among several. Viewpoint space was constrained to two dimensions.

6.2.2. Octree Methods.

Octree methods encode voxel occupancy more efficiently.

Connolly. One of the earliest papers on view planning was by Connolly [1985]. He appears to have first coined the term "next-best-view" (NBV). Connolly presented two algorithms, planetarium and normal, differing in the cue used to suggest feasible views and the selection mechanism used to find the next-best-view from the set of feasible candidates. The imaging work space is voxelized and labeled as empty, occupied, or unseen. This information is encoded in an octree which is also used to represent the object surface. All viewing vectors are assumed to point to the origin. Viewpoints are constrained to evenly spaced points on a sphere around the object.

The planetarium algorithm applies a visibility analysis of the octree for each candidate viewpoint on the sphere. The area of unseen voxels projected onto the image plane for each candidate viewpoint is taken as a measure of the solid angle of unseen space that will be swept by that view. The viewpoint with the largest unseen area is selected as the NBV. The algorithm suffers from time complexity problems inherent with a complete visibility analysis for all candidate viewpoints. The normal algorithm simplifies the visibility analysis to the local rather than global level by examining faces in the octree common to both unseen and empty voxels. It is faster but does not deal as well with self-occluding scenes.

The method is subject to many simplifications and limitations. Neither sensor nor positioning system performance is characterized. The sensor is treated as an ideal point source without constraints on field-of-view or depth-of-field. Shadow effects are ignored. Notwithstanding these limitations, Connolly's pioneering concept of exploring and labeling imaging work space is found in many later papers.

6.2.3. Space Carving.

The space carving technique[6] applies primarily to a small class of range sensors with a limited sensing volume, particularly those designed as noncontact replacements for CMM mechanical touch probes. The sensor is swept through the imaging work space in a preplanned methodical manner, diverting around obstacles, with the objective of

---

[6] As used here, "space carving" is distinct from the shape from silhouette technique [Kutulakos and Seitz 2000].

reliably labeling work space occupancy. As with contour following, collision avoidance is a primary concern.

Papadopoulos-Orfanos. A paper by Papadopoulos-Orfanos and Schmitt [1997] summarizing Papadopoulos-Orfanos' [1997] Ph.D. research applied space carving to a specific shallow depth-of-field sensor for automated object reconstruction. Emphasis was placed on updating the voxel-based scene representation—in particular, robust labeling of empty space to avoid collisions. Voxel labeling was achieved by a complex analysis of occlusion zones followed by ray tracing operations. Most of the work was concentrated on collision avoidance and path planning.

Papadopoulos-Orfanos described but did not fully implement a two-stage 3D imaging strategy of scene exploration followed by surface data acquisition. The exploration stage, which was implemented, employed an exhaustive search of unknown portions of the work space. In the vicinity of the object, the goal was to get the sensor as close as possible to the surface while avoiding collisions. During this phase, sensor orientation was fixed and only translations were employed to search the work space layer by layer in a zigzag pattern. As a consequence of the fixed orientation, occluded surfaces were not imaged and surface occupancy in these regions was poorly defined. In a second stage of surface data acquisition (described but not implemented), sensor orientation view planning was proposed. Some existing NBV techniques were briefly discussed as candidate methods.

In its current form, the work does not address view planning per se. The exploration stage involves an exhaustive search of the imaging work space following a preplanned trajectory modified by collision avoidance. No new view planning technique is proposed. If more fully developed, space carving has potential for high-precision scanning. However, it will likely remain slow as a consequence of the small sensor frustum and the exhaustive search technique.

Lamb et al. [1999] also utilized space carving at a coarse level of resolution in the first phase of a multistage approach to semiautomated model acquisition.

6.2.4. Solid Geometry Methods.

This method utilizes standard solid geometry algorithms available with most CAD packages to model the current state of object knowledge. The method can be robust with respect to complex topology. A generic problem with the technique arises from solid geometry intersection operations which, by definition, subtract and cannot add volume. Therefore, if a range image used to extrude a solid volume does not completely cover the object, the resulting volume will exclude a portion of the object which can never be recovered by subsequent intersection operations. Consequently, difficulties arise when a view fails to completely enclose the object or along occlusion boundaries where data is often missing or erroneous due to inclination and edge effects.

Bistatic shadow effects. Although ignored by many authors, a triangulation-based laser scanner is a bistatic sensor whose optical baseline is generally significant with respect to the measurement stand-off distance. Therefore, the shadow effect is nonnegligible and has two components—source and receiver shadows. Bistatic shadow effects are illustrated in Figure 6. We can observe that an occlusion edge as seen by the receiver may lie inside the object. This poses difficulties for occlusion-edge-based view planning techniques employing solid geometry intersection operations. Collision avoidance routines operating in conjunction with view planning also need to make allowance for this fact. Further, in general, the object will not lie entirely within the camera frustum, so a given range image may or may not contain boundary occlusion edges. Some NBV algorithms can fail in the absence of occlusion edges.

Reed. In a series of articles by Reed and Allen [1997, 1999] and Reed et al. [1997a, 1997b] arising from Reed's [1998] Ph.D.

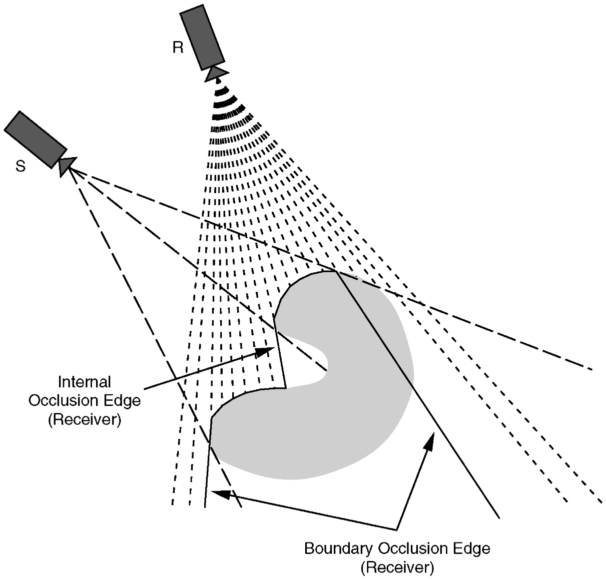

Fig. 6. Occluding edges in geometric images with bistatic sensor.

thesis, a solid geometry approach was used to synthesize a continuous 3D viewing volume for a selected occlusion edge, rather than discretizing viewpoint space. A mesh surface was created for each range image and each mesh triangle was swept to form a solid model. Solid geometry union operations on the swept volume resulted in a composite model comprised of the surface and occluded volume, as known at that stage of modeling. Surfaces on the composite model were labeled *imaged surface* or *occluded surface*.

Reed's view planning approach then proceeded with manual selection of a specific occluded edge as a target. A "visibility volume" was calculated for the target—that is, a 3D volume specifying the set of all sensor positions having an unoccluded view of the entire target for the model as presently defined. This was computed by subtracting the occlusion volume for each model surface component from the target's unoccluded volume, a half-space whose defining plane was coincident with the target's face and oriented in the appropriate direction. The work was based on previous sensor planning work in the same lab by Tarabanis et al. [1996]. The volumetric approach allowed sensor and positioning system constraints to be added.

The 3D imaging workspace search was an improvement over methods constrained to an arbitrary 2D surface around the object. The authors claimed their approach had the advantage of defining more accurate viewpoints due to the treatment of viewpoint space as continuous rather than discrete, although no analytical or experimental data were provided to quantify this assertion.

Part of the reported research [Reed et al. 1997a] relied on manual selection of a suitable target and hence the NBV. There was no discussion of the sensitivity of the algorithm to target occlusion edge selection. A second paper [Reed et al. 1997b] discussed an approach to automated viewpoint selection. In this case, the algorithm selected the NBV as the one imaging the most occluded surface elements in the current composite model. The NBV was determined by computing the visibility volume for each occluded surface element, intersecting each visibility volume with the sensor's reachable space and searching the intersection for the point imaging the most surface area. A third paper [Reed and Allen 1999] briefly examined NBV planning based on the use of multiple targets.

Benefits of solid geometry methods include robustness to complex object topology and the guarantee of watertight models. Viewpoint synthesis can be computationally advantageous with respect to methods which discretize viewpoint space. However, as presently defined, the approach has limitations. The sensor frustum is only partially modeled—field-of-view constraints are not considered. Bistatic shadow effects on view planning are not addressed directly, although missing data in small shadow zones is dealt with by interpolation and along occlusion boundaries by surface extension along the direction of the sensor baseline. Sensor errors are only briefly considered. View overlap planning is not addressed. Set operations on extruded surface elements introduce artifacts along integration boundaries.

6.3. Global View Planning Methods

A few methods derive a view planning cue from global rather than local characteristics of the geometric data.

6.3.1. Mass Vector Chain. Yuan [1993, 1995] described an interesting view planning mechanism. He observed that "a reconstruction system expects a self-controlled modeling mechanism for the system to check the spatial closure of object models and to estimate the direction of unprocessed features" [Yuan 1995, p. 307].

Illustrated in Figure 7, his approach segments the observed surface into a set of small patches and describes the object with a mass vector chain (MVC). By definition, an MVC is a series of weighted vectors. A mass vector $\vec{V}_i$ is assigned to each surface patch S_i of the object. Thus, $\vec{V}_i = \vec{n}_i R_i$, where $\vec{n}_i$ is the average visible direction and R_i is the surface size when viewed from that direction. It is easily shown that the boundary surfaces of an object compose a closed surface boundary only when their mass vectors form a closed chain. Thus, the next-best-view can be set to the negative of the cumulative MVC, which should define a viewing direction whose mass vector will close the chain or at least shorten the gap. Special rules are required for holes and cavities.

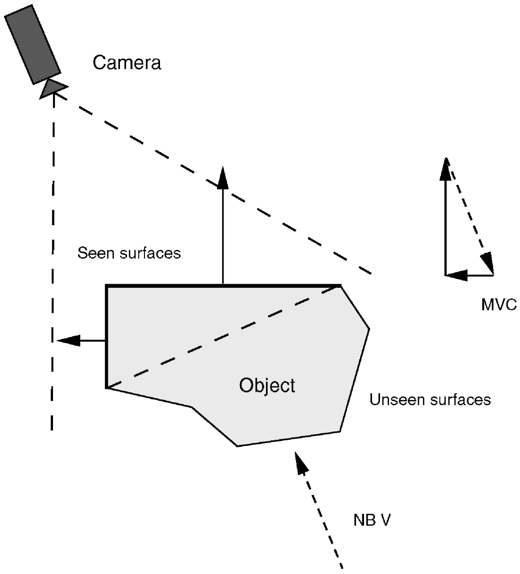

Fig. 7. Mass vector chain.

The algorithm was shown to work with synthetic geometric data for a simple object. However, the idealized imaging environment does not address the complications of modeling automation with real sensors and complex humanly made or natural objects. The method is capable only of estimating viewing direction, not position. It has no reliable measure of scale, range, or imaging volume. The method is able to deal with moderately complex topology and may have potential for machine vision tasks involving humanly made parts with simple shape or for the initial global view planning phase of a two-step coarse-fine planning process.

6.3.2. Intermediate Space Representations. The essence of view planning is capturing, representing, and optimizing measures of visibility of the object surface from sensor poses in viewpoint space. This mapping between the object surface and points in the workspace of the sensor and positioning system involves a large amount of information and therefore a high degree of computational complexity for its acquisition, storage, and manipulation. It is desirable to find a more compact and easily searched representation of the visibility information with a minimum loss of information. One approach following this line of reasoning involves encoding visibility information on a virtual surface positioned between the object and the sensor workspace—that is, an *intermediate space representation*.

Positional space—Pito. Pito's [1997a] thesis is significant for its contributions to all four aspects of geometric modeling automation (scanning, registration, integration, and view planning). See also the summary paper [Pito 1999] as well as Pito [1996a, 1996b, 1997b] and Pito and Bajcsy [1995].

Pito's work is unusual in that he took pains to characterize performance of the range scanner used for experiments and then incorporated sensor models in the automation algorithms. In addition to treating the grazing angle effect (fairly common in the literature), he also treated edge effect anomalies during the integration phase. The view overlap requirement was incorporated. The shadow effect was explicitly treated. However, nonstationary measurement error within the measurement volume was not addressed. The need for generalized viewpoints was discussed but apparently not implemented. The technique is amenable to using generalized viewpoints, at the expense of increasing the dimensionality of intermediate space.

Like many others before him, Pito chose occlusion edges as a cuing mechanism for the NBV search. He observed that the void volume can be economically represented by defining only the void surface near edges of the current model, which he represented by small rectangular patches attached to occluding edges of the seen surface. He argued that it is unnecessary to represent the complete boundary of the umbra cast by the object. In general, the portion of the umbra nearest the occlusion boundary will be nearest to the real object surface and therefore be the best region to search next. Furthermore, such regions fit the overlap constraint.

Illustrated in Figure 8, Pito's NBV algorithm used an intermediate space representation, "positional space" (PS), as a repository for two types of visibility information—object surface visibility and sensor scanning potential. A virtual positional space surface (PSS), with a shape appropriate to the sensor-positioning system combination, was placed between the object and the sensor workspace. Object surface space (represented as a mesh), PSS, and viewpoint space were discretized.

The visibility of each triangular mesh element on the object surface can be encoded in positional space by tracing an "observation ray" from the mesh element to a PSS cell. The direction of an unoccluded observation ray relative to a local frame of reference on the PSS can be specified by two angles, termed the *positional space direction* (PSD). Pito chose to encode surface visibility as an analogue value equal to the area of the surface element visible by a given ray weighted by a confidence measure tied to the measurement grazing angle. Thus, PS was a scalar field in four dimensions $P(u, v, \theta, \phi)$ where u, v were coordinates in PSS and θ, ϕ were the components of PSD. Encoding the image of both the seen surface and void patches in PS provided a means to apply an overlap constraint between views to meet registration and integration requirements. The range camera's scanning potential at a given viewpoint can be similarly encoded in positional space by determining the intersection of "ranging rays" from the optical transmitter and receiver with the PSS. A separate image was calculated for each viewpoint. Using PS as a placeholder for ranging and observation rays facilitated determining which of them were collinear, as well as aiding the application of NBV constraints.

Without detracting from the comprehensive treatment and contribution of the work, there are some unresolved issues

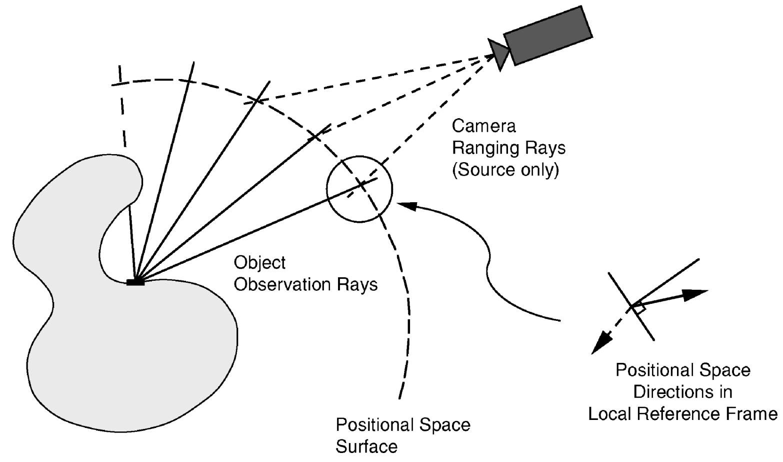

Fig. 8. Positional space.

and weaknesses with the approach:

—Overall computational complexity issues are not addressed; nor are tradeoffs between discretization levels in object surface space S, viewpoint space V, or positional space PSS. In the end, visibility analysis relates S to V. The impact of quantization errors arising from imposition of an intermediate space between S and V is not addressed. Efficiency is a concern. The experimental setup used to demonstrate the algorithm was limited to a single rotational degree of freedom.

—One of the major claims of the work was that the computational burden of visibility analysis is decoupled from the size of viewpoint space. Nevertheless, ray tracing figures heavily in the approach and it has not been demonstrated that the net visibility analysis burden is lowered for a given performance level. However, the algorithm appears to be readily parallelizable.

—PS representation is a lossy compression of pose information in which range-related information is lost. Stand-off range critically impacts sensor performance.

—A PS image must be cast for each viewpoint, a potentially large number. The algorithm does not provide guidance on efficiently sampling viewpoint space. On the other hand, PS sensor images need be calculated only once and stored offline. Memory issues are not addressed.

—Reference is made to the use of multiple PS surfaces which would capture additional information in viewpoint space but this would appear to reduce the claimed computational advantages of the representation.

6.3.3. Expert System. Artificial intelligence and expert system approaches to view planning have been few and limited mainly to illumination issues [Batchelor 1989; Novini 1988; Kitamura et al. 1990].

7. COMPARISON OF EXISTING METHODS

Recall the following (for papers being referenced, see discussion in Section 6):

—*Maver and Bajscy* selected an NBV by reasoning about the angular arc through which occluded areas of the object surface can be viewed, based on assumptions about the unseen topography.

—*Connolly* selected an NBV as the one able to acquire the most information about the unseen imaging volume, based on either a global or local visibility analysis.

- *Yuan* selected an next-best-viewing direction based on a global analysis of the mass vector chain for the object surface.
- *Banta et al.* discussed several view planning mechanisms, the most recent being the NBV oriented to the centroid of the cluster containing the largest number of unknown voxel faces.
- *Massios and Fisher* used an objective function incorporating both visibility and "quality" terms.
- *Tarbox and Gottschlich* computed a measurability matrix encoding a complete visibility analysis over the set of all viewpoints and surface elements, and selected a viewpoint maximizing the void surface visible.
- *Reed et al.* first synthesized a visibility volume for a selected occluded edge using solid geometry techniques and then selected an NBV imaging the most occluded surface elements.
- *Whaite and Ferrie* used uncertainty in fitting parameterized superquadrics to segmented range data as a cue to viewpoint selection and then selected a viewpoint at a fixed offset in the direction in which the variance of the fit of the data to the current model is the greatest.
- *Pito* separated the visibility analysis of object and sensor by casting visibility images onto an intermediate space representation, positional space, and optimized viewing direction selection within that domain subject to various constraints.
- *Papadopoulos-Orfanos* defined a space carving technique for a shallow depth-of-field range sensor based on a preplanned search pattern plus collision avoidance.

Table III summarizes attributes of view planning techniques with respect to requirements specified in the introduction. The legend for the comparison table is as follows: requirement satisfied (Y), not satisfied (N), partially satisfied (P), uncertain (?), or not applicable (−). The objective is a column of all Y's.

8. CRITIQUE OF EXISTING METHODS

8.1. General

8.1.1. Model Quality Specification. No method reviewed incorporates explicit quality goals for the reconstructed object model. Massios used a quality term in the view planning objective function but this term was not tied to explicit pass/fail criteria. Several authors included a "measurement quality" term tied to grazing angle but did not relate the factor to explicit model performance requirements.

8.1.2. Generalizable Algorithm. Several methods are generalizable to a broad class of range sensors, positioning systems and objects. Others are not.

8.1.3. Generalized Viewpoint. Several authors described a generalized viewpoint in their opening theoretical problem formulation. However, none retained the formulation during algorithm development. In most cases, pose space was constrained to one or two dimensions versus the generalized case which is $6D^+$. No author considered a programmable sensor parameter other than pose. Some methods (Yuan, Pito) are capable of providing only incomplete viewpoint information, specifically viewing direction but not sensor position or orientation about the sensor boresight.

8.1.4. View Overlap. Almost all authors assumed error-free positioning and ignored view planning constraints on view overlap for image-based registration. Only Pito applied a view overlap constraint, although shape complexity was not addressed.

8.1.5. Robust. None of the methods incorporated both sensor and positioning system error models. Several methods (Massios, Tarbox, Pito) considered grazing angle effects as a subjective quality factor. No method considered geometric step edges, reflectance step edges, or multiple reflection effects. The robustness of the proposed methods in realistic sensing and view planning environments is either

Table III. Comparison of View Planning Algorithms

Category	Requirement	Maver & Bajscy	Connolly	Yuan	Banta et al.	Massios & Fisher	Tarbox & Gottschlich	Reed et al.	Whaite & Ferrie	Pito	Papadopoulos-Orfanos
General	Model quality specification	N	N	N	N	N	N	N	N	N	N
	Generalizable algorithm	N	Y	N	Y	Y	Y	Y	N	Y	N
	Generalized viewpoints	N	N	N	N	N	N	N	N	N	N
	View overlap	N	N	N	N	N	N	N	P	P	N
	Robust	N	N	N	N	?	?	?	N	?	?
	Efficient	N	?	?	?	?	N	?	N	?	N
	Self-terminating	N	Y	Y	Y	Y	Y	Y	Y	Y	Y
Object	Minimal a priori knowledge	Y	Y	Y	Y	Y	—	Y	N	Y	Y
	Shape constraints	N	Y	N	Y	Y	Y	Y	N	Y	Y
	Material constraints	N	N	N	N	N	N	N	N	N	N
Sensor	Frustum modeled	N	N	N	N	N	N	P	N	P	Y
	Shadow effect	Y	N	N	N	N	Y	N	N	Y	Y
	Measurement performance	N	N	N	N	P	P	N	P	P	N
Positioning System	6D pose	N	N	N	N	N	N	N	N	N	Y
	Pose constraints	N	N	N	N	Y	Y	Y	N	Y	Y
	Positioning performance	N	N	N	N	N	P	N	N	N	N

Note: For papers being referenced, see discussion in Section 6.

weak or is undetermined—that is, the environment of real sensors with real errors and artifacts, real objects with representative topological and geometrical complexity, and real positioning systems subject to pose error and restricted movement envelopes.

8.1.6. Efficiency. Given the lack of standardized performance benchmarks, it is difficult to make quantitative performance comparisons between proposed view planning methods. However, some authors (Banta, Tarbox, Connolly) described multiple algorithms and provided qualitative performance comparisons. Few authors explicitly evaluated the computational complexity of their algorithms. Computational complexity is an issue with all methods. However, many of the techniques are parallelizable and computational power and memory are becoming less of a concern with ongoing advances in computer technology.

8.1.7. Self-Terminating. Various termination criteria were used, but none related to explicit requirements to be met by the reconstructed object model.

8.2. Object Characterization

8.2.1. Minimal a priori Knowledge. Most methods imposed no a priori knowledge constraints on the objects to be modeled. Tarbox's work focused on inspection. Whaite and Ferrie's approach was aimed at robotic environment exploration.

8.2.2. Shape Constraints. Most methods imposed no object shape constraints. However, some would clearly be unable to handle complex shapes due the coarseness of the underlying approximations, for example, the $2^1/_2$ D scene view of Maver and Bajcsy [1990], the mass vector chain approach [Yuan 1993], or parametric surface representation [Whaite and Ferrie 1997]. Performance was not characterized with respect to increasing object topological or geometrical complexity.

8.2.3. Material Constraints. No method dealt with object material effects on sensor performance.

8.3. Sensor Characterization

8.3.1. Frustum Modeled. Most methods did not incorporate a realistic model of the sensor frustum, that is, the depth-of-field, angular field-of-view, and scan length/arc of real range cameras.

8.3.2. Shadow Effect. Few methods modeled shadow effects. This core attribute of triangulation-based range sensors limits observation in cavities or around obstacles. A great deal of manual view planning effort is spent addressing shadow zones. High-performance triangulation-based range sensors can be designed with small optical baselines, for example, "BIRIS" and autosynchronous scanners (www.vit.iit.nrc.ca). Small bistatic shadow effects benefit some view planning techniques, notably solid geometry methods as well as all occlusion edge methods.

8.3.3. Measurement Performance Characterization. With some exceptions [Pito 1999], the view planning literature generally assumes idealized, error-free sensors. Many model the camera as an ideal monostatic sensor. Yet geometric measurement techniques are highly nonlinear and error prone. A substantial portion of the effort in manual object reconstruction is spent minimizing errors in data acquisition and model building. The sensing error most commonly considered in the literature is inclination bias. No author addressed the nonisotropic and nonstationary nature of residual geometric noise within the sensor frustum of a calibrated range sensor. Traditional view planning approaches have attempted to obtain complete coverage of the object surface with limited attention to the quality of the geometric data.

8.4. Positioning System Characterization

8.4.1. 6D Pose. Most methods limit viewpoint space to one or two dimensions. The most common configuration constrains the sensor to a "viewing sphere" with the camera boresight fixed on an object-centered origin. It is not possible to optimize sensor performance with such limitations.

8.4.2. Pose Constraints. Several of the more advanced approaches to view planning incorporated positioning system constraints.

8.4.3. Positioning Performance. Excepting Tarbox, no author addressed the impact of pose error on view planning. Tarbox removed viewpoints not robust to pose variation from the viewpoint set associated with each surface point. Positioning system performance was not modeled.

8.5. Critique Summary

8.5.1. Techniques and Representations. By far the most common NBV cuing mechanism is either the surface or the solid volume of the umbra cast by a range image view of the scene—that is, occlusion edges or the shadow zone. The greatest variability between view planning approaches lies in the mechanism used to select the NBV. Common surface representations are volumetric occupancy and surface meshes. The most common representation of imaging work space is volumetric—voxel occupancy, octree, or solid geometry.

8.5.2. Object Size Assumption. Many techniques assume that the object falls completely within the camera field of view and would fail if this condition were not met. For example, if a range image contains no jump edges, what should the algorithm do next?

8.5.3. Strategies. Most approaches are monolithic in the sense that a single technique was used throughout. It would appear advantageous to divide the overall view planning problem into stages as Hall-Holt [1998] suggested and consider different techniques at different stages—in particular, restricting computationally intensive techniques to limited subsets of the overall problem. All of the methods examined are deterministic. There may be opportunities for random selection methodologies.

8.5.4. Performance Objectives and Evaluation. No current method includes explicit pass/fail quality criteria on the outcome. Generally, there is a lack of clarity with respect to the objective. Is view planning a search for global optimality or merely acceptable viewpoints? What determines success—a minimum number of viewpoints; time to complete the job; some other criteria?

8.5.5. Suitability. As presently developed, no traditional view planning method is directly suitable for performance-oriented automated object reconstruction for the following principle reasons: overly constrained viewpoint space, inadequate sensor and positioning system performance characterization, and excessive computational complexity.

9. OTHER VIEW PLANNING ISSUES

9.1. Theoretical Framework

Theoretical treatment of the view planning problem has been limited. Tarbox and Gottschlich [1995] introduced the measurability matrix concept in a model-based approach to inspection. They showed the VPP to be isomorphic to the set covering problem which is known to be NP-complete. Yuan [1995] used mass vector chains in a global approach to view planning. Whaite and Ferrie [1997] presented an autonomous exploration theory based on minimization of model uncertainty. Soucy et al. [1998] examined view planning from the perspective of an infinitely small surface explorer and local topological operations on a voxelized surface. They also took the capabilities and limitations of the positioning system into account. Arbel and Ferrie [1999] presented an entropy-based gaze planning strategy in the somewhat related field of object recognition. Scott et al. [2000] extended Tarbox's work to object reconstruction with a set theory-based formulation of the VPP in terms of a measurability mapping between viewpoint space and object surface space. Scott et al. [2001a] later showed that the view planning problem can be expressed as the problem of covering the rows of a binary measurability matrix by a minimal subset of its columns and provided an integer programming formulation of the problem with a registration constraint. Most other work in the field has relied on a variety of heuristic techniques without a theoretical framework.

The referenced work has contributed to a better understanding of the view planning problem, yet the field lacks a comprehensive theoretical foundation. A truly comprehensive theory would encompass all problem constraints and error mechanisms and provide a theoretical basis for sampling surface space and viewpoint space.

9.2. Viewpoint Generation

The view planning process computes a set of viewpoints to satisfy the specified reconstruction or inspection goal. This is usually done by synthesis or generate and test methods. Synthesis methods (Tarabanis et al. [1995b]; Reed et al. [1997a]) are in the minority. In principle, synthesis methods have the advantage of simultaneously optimizing competing factors over multidimensional viewpoint space but, in practice, computational difficulties of nonlinearity and convergence usually arise. The great majority of view planning methods follow the generate and test paradigm—that is, viewpoint space is discretized by some method and the view planning algorithm selects a set of these by some optimization algorithm.

9.2.1. The Concept of a "Viewpoint."

Some subtleties in use of the term *viewpoint* merit elaboration. In this survey, we consider a viewpoint to define a 3D frustum. However, most triangulation range cameras are actually 2D sensors. They measure depth z over a 2D fan-shaped region swept by the laser in the sensor x-z plane. The third measurement dimension is achieved by physically moving the sensor perpendicular to the plane of light. This can be achieved by a sweeping motion along the camera y-axis (line-scan mode) or rotation about an axis parallel to the camera x-axis (cylindrical-scan mode). Range data gathered in such a manner will form a "range image" in the conventional sense—that is, a u-v grid of depth pixels.

A different data acquisition approach is to move the sensor through an arbitrary arc in space to collect a long swath of laser scan profiles. The motion velocity vector need not have constant speed or direction. Sensor orientation may be constant or variable during the sweep. Lamb et al. [1999] and Soucy et al. [1998] employed this mode in contour following strategies using a CMM as the positioning system. In their setup, sensor position could be changed readily but orientation changes were costly in reconfiguration and recalibration time. Space carving techniques such as those used by Papadopoulos-Orfanos and Schmitt [1997] acquire range data in a similar manner.

Our use of the term *viewpoint* follows the first of these conventions. A viewpoint defines a single discrete position and orientation in space, with an associated constant 3D frustum shape. In addition to pose, a generalized viewpoint also has a set of configurable sensor parameters. We associate "range images" with viewpoints. For most cameras, range image size (number of range pixels) is either fixed or is programmable over a narrow range of values.

A more appropriate term for the sensing operation commonly used for contour following and space carving would be a *view swath*. A view swath has a constant 2D frustum defined in the scanning plane. A view swath defines a continuous set of viewing positions and orientations over some arc in space, in effect a collection of 2D laser profiles. Neither speed nor direction of the velocity vector defining the sweep arc need be constant. View swathes are associated with either a set of 2D range profiles of arbitrary length or a 3D point cloud of arbitrary size.

9.2.2. Generalized Viewpoints.

Most view planning methods sharply constrain the dimensionality of viewpoint space to limit computational complexity. Few authors have recognized the importance of sensor parameters in addition to pose. Tarabanis [1995b] and Pito [1999] were exceptions. View planning for high-performance sensors and difficult reconstruction tasks cannot ignore configurable sensor parameters.

9.2.3. View Sphere.

The most common stratagem for constraining viewpoint space V uses a virtual, object-centered "view sphere" enclosing the object ([Banta and Abidi 1996a; Connolly 1985a; Garcia et al. 1998a; Massios and Fisher 1998; Morooka et al. 1999; Reed et al. 1997b; Tarbox and Gottschlich 1995; Yi et al. 1995]). This reduces the view planning problem to the task of finding the best set of viewing directions. The viewing axis of any viewpoint on the sphere is oriented to the center of the reference frame. Variable orientations about the viewing axis have rarely been considered. This stratagem reduces the dimensionality of V from $6D^+$ to 2D. A number of authors have gone further, reducing V to a circular 1D domain, a simplification trivializing the problem and masking the underlying complexity. Many authors also appear to have treated triangulation-based range cameras as monostatic devices, thus dismissing shadow-effects, an intrinsic sensor characteristic and key aspect of realistic view planning.

A few synthesis methods ([Reed et al. 1997b; Tarabanis et al. 1995b]) have treated the view sphere as continuous. The majority, however, have used discretization. This includes parallels and meridians [Connolly 1985; Zha et al. 1997], spherical distribution maps (SDMs) ([Garcia et al. 1998a; Ye and Tsotsos 1999; Yi et al. 1995]) or, the most popular, a tessellated icosahedron ([Massios and Fisher 1998; Morooka et al. 1999; Sakane et al. 1992; Tarbox and Gottschlich 1995; Trucco et al. 1997; Zha et al. 1998]). While it is well known that a uniform tessellation of a sphere does not exist, a tessellated icosahedron provides a close approximation. Morooka et al. [1999] went further, transforming the geodesic tessellation into a flattened spherical array to improve efficiency in mapping orientation to geodesic cells.

Discretization by parallels and meridians offers simplicity at the expense of nonuniformity. A tessellated icosahedron provides an almost uniform subdivision but may require complex computations regarding cell occupancy. The SDM approach offers intermediate performance—fast mapping of orientations to cells with reasonable cell uniformity, although inferior to the tessellated icosahedron.

Visibility analysis is an essential but computationally costly component of many view planning algorithms. The approach of Tarbox and Gottschlich [1995] is unique in that viewpoint space was defined as the set of all points on the view sphere separated by a distance equal to the sensor optical baseline. For a tessellated icosahedron, this provides five or six orientations about the sensor boresight per viewing axis. In Tarbox's algorithm, view sphere vertices serve as locations for either the optical source or receiver. Ray tracing is conducted once per vertex rather than twice per viewpoint, thus reducing the cost of visibility analysis.

Almost all users of the view sphere strategy select a fixed stand-off distance such that the sensor frustum encloses the whole object. This may be necessary for some conventional surface-based techniques relying on occlusion edges to cue the next-best-view selection scheme. However, the strategy fails to take into account the important dependence of sensor performance on stand-off distance, a key feature of all triangulation-based range cameras. Also, it fails completely for imaging tasks for which object size exceeds frustum dimensions.

The principal advantages of the view sphere technique are twofold: a reduction in the dimensionality of the pose component of viewpoint space from six to two, and relatively uniform sampling of the axis component of orientation (except for discretization by parallels and meridians). Its main disadvantages are as follows. To the degree that object shape is not spherical, viewpoint position is nonoptimal. Measurement performance is further degraded if the view sphere radius must be set to view the entire object. Additionally, the view sphere ties orientation to position whereas there is more to achieving good viewing directions than merely selecting the orientation of the sensor boresight. The position from which the image is taken is also important. Further, the view sphere approach says nothing about the orientation of the camera around its boresight (the twist angle). For mid- to high-performance range cameras, the optical baseline typically subtends an angle in the range (50°–35°) at the sensor's optimal stand-off range. Consequently, the shadow effect is sensitive to rotation about the sensor boresight.

9.2.4. Intermediate Space. Pito [1999] separated visibility analysis of the object surface from that of the sensor by casting visibility images onto a virtual intermediate space representation, *positional space* (PS), positioned between the object's convex hull and the sensor. Visibility is verified by comparing the orientation of sensor ranging rays and object observation rays as encoded with respect to the local reference frame in positional space. However, PS representation is a lossy compression of pose information in which range-related information is lost. Stand-off range critically impacts sensor performance.

9.2.5. Viewpoint Generation Summary. In summary, most techniques constrain viewpoint space to two dimensions, a view sphere or view cylinder, and in some cases to a one-dimensional viewing circle. Viewpoint positions are constrained to this surface or arc and the sensor boresight is fixed on the object center. Orientation about the sensor boresight is fixed (excepting the approach by Tarbox). The view sphere radius is also fixed, most commonly to a distance ensuring visibility of the complete object. The requirement to consider generalized viewpoints has not been widely recognized. The view sphere approach is mathematically attractive and provides reasonably uniform sampling in orientation but fails to optimize position or some aspects of orientation. Finally, the practicality of matching virtual spherical viewpoint domains with the capabilities and limitations of real positioning systems, real sensors, and real objects is rarely considered.

9.3. Pose Error

View planning is a computationally intensive task with the objective of arriving at a small set of optimal or near-optimal viewpoints. When the NBV list is sent to a positioning system whose position and orientation accuracy is inferior to that of the sensor, the coverage of individual viewpoints and of the NBV list as a whole is compromised. Individual viewpoint positions and orientations are corrupted. Orientation error is particularly troublesome as effects are amplified by range. As illustrated in Figure 9, image coverage (frustum occupancy), measurement precision, sampling density, and visibility will all be effected.

We can recover a refined pose estimate post facto by employing suitable registration techniques and subsequently reestimate measurement quality within the acquired image. However, we are still left with data acquisition differing from that which had been planned. As pose error deteriorates, the computationally intensive view planning phase is progressively compromised—ultimately to be rendered futile. Consequently, there is a need to make the view planning process robust with respect to pose uncertainty resulting from positioning system errors.

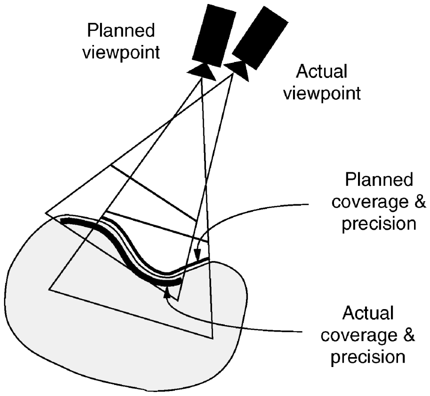

Fig. 9. View planning with pose uncertainty.

The problem of pose error has received little attention in the view planning literature. Tarabanis et al. [1995b] used a synthesis approach for generalized viewpoints which seeks to centralize viewpoints in the admissible domain. Tarbox and Gottschlich [1995] used morphological erosion of viewpoints on the periphery of viewpoint sets to reduce vulnerability to pose error. Recently, Scott et al. [2002] examined pose error effects on range sensing.

10. CONCLUSION

10.1. Open Problems

This paper has surveyed and compared view planning techniques for automated 3D object reconstruction and inspection by means of active, triangulation-based range sensors. Subject to previously noted strengths and weaknesses, some current methods are usable in a production setting if suitably integrated and if reconstruction/inspection quality requirements are relaxed, particularly for a scene exploration stage. This would include voxel occupancy, occlusion edge, space carving, solid geometry, contour following, and intermediate space techniques. These comments apply to objects with relatively simple shape and benign reflectance properties. As yet, view planning for high-quality object reconstruction/inspection has no general-purpose solution. There are no known general-purpose automated geometric modeling environments or performance-oriented view planning systems in public usage. The remaining open problems are simultaneously both intriguingly simple and exceptionally complex—efficiency, accuracy, and robustness.

The efficiency issue relates to the computational complexity of the view planning algorithm in terms of both time and memory, although timeliness is probably the more important factor. No method proposed to date is able to provide both performance and efficiency. In fact, even with unrealistic constraints on the problem domain, most view planning techniques are unacceptably slow. However, ongoing computer technology developments are making this less of a concern.

Accuracy and robustness considerations go hand-in-hand. Both demand that the view planning module incorporate reasonably complete sensor noise and artifact models based on performance characterization of the sensor and positioning system. Additionally, high-accuracy object reconstruction requires a highly mobile positioning system. This means desirably a full six degrees of freedom over an adequate imaging work space. View planning algorithms and imaging environments restricted to a one- or two- dimensional viewpoint space can at best provide limited-quality, limited-coverage models for a restricted class of objects.

Finally, there is a need to measure and model object reflectance to handle shiny surfaces and compensate for limited range camera dynamic range. View planning techniques are also required to avoid or compensate for the effects of geometric step edges, reflectance step edges, and multiple reflections.

10.2. Future Research Directions

This survey of current view planning techniques suggests more flexible, multifaceted methods:

1. a multi-phase approach such as scene exploration, rough modeling, fine modeling, problem resolution;
2. use of multiple techniques in combination instead of a monolithic approach using a single technique;
3. heuristic and expert system designs emulating human operators; and
4. multisensor fusion approaches—perhaps using a fast, wide field-of-view, low-precision sensor for scene exploration and collision avoidance in combination with a high-quality range sensor for precise surface measurements.

Overcoming dropouts and outliers due to limited sensor dynamic range remains an important open issue. Part of the solution will be found in sensor improvements, in particular digital signal processor-based advanced signal processing on the raw data. Further improvements should be possible by incorporating reflectance measurement and modeling in a scene exploration stage, which could also address two other important error sources—reflectance step edges and multiple reflections.

Additionally, it is evident that the field could benefit from standardized performance benchmarks and quantitative algorithm performance analysis. Finally, to date there has been only limited treatment of theoretical aspects of the problem.

Predicting future developments is hazardous but we will make a few projections. The growing capability of low-cost computers will mitigate the previously prohibitive computational expense of visibility analysis, a core element of many view planning techniques. Practical view planning solutions are unlikely to be monolithic but rather to involve multiple stages, techniques, and sensor types. Model-based approaches are likely to be fruitful, particularly for high-quality reconstruction/inspection applications. Positioning system performance and cost will continue to be a problem. This is an understudied area. Technology and applications for 3D sensors will continue to proliferate rapidly, mainly in the directions of speed and lower cost, and more slowly in performance. Cost, time, and user skill level issues will grow, increasing pressure for automation solutions. Dynamic range limitations of optical sensors will continue to present challenges for the foreseeable future. Fully automated object reconstruction/inspection is some distance off and will involve not just research but complex system engineering and integration challenges. Nearer-term solutions will provide semiautomated aids to skilled operators.

REFERENCES

Abrams, S. 1997. Sensor planning in an active robot work-cell. Ph.D. dissertation, Columbia University, New York, NY.

Amann, M.-C., Bosch, T., Myllyla, R., and Rioux, M. 2001. Laser ranging: a critical review of usual techniques for distance measurement. *Opt. Eng. 40*, 1, 10–19.

Arbel, T. and Ferrie, F. 1999. Entropy-based gaze planning. In *Proceedings of the 2nd IEEE Workshop on Perception for Mobile Agents, in association with 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Fort Collins, CO.). 87–94.

Aubry, S. 1994. 3D model construction from multiple sensor viewpoints. Ph.D. dissertation, McGill University, Montreal, P.Q., Canada.

Banta, J. and Abidi, M. 1996. Autonomous placement of a range sensor for acquisition of optimal 3D models. In *Proceedings of the IEEE 22nd International Conference on Industrial Electronics, Control and Instrumentation* (Taipei, Taiwan). 1583–1588.

Banta, J., Zhien, Y., Wang, X., Zhang, G., Smith, M., and Abidi, M. 1995. A best-next-view algorithm for three-dimensional scene reconstruction using range images. *SPIE 2588*, 418–429.

Banta, J. E. 1996. Optimal range sensor positioning for three-dimensional model reconstruction. M.S. thesis, University of Tennessee, Knoxville, TN.

Baribeau, R., Rioux, M., and Godin, G. 1991. Color reflectance modeling using a polychromatic laser range sensor. *IEEE Trans. PAMI 14*, 2 (Feb.), 263–269.

Batchelor, B. 1989. A prolog lighting advisor. In *Proceedings of the SPIE Conference on Intelligent Robots and Computer Vision VIII: Systems and Applications*. Vol. 1193. 295–302.

Beraldin, J.-A., El-Hakim, S., and Cournoyer, L. 1993. Practical range camera calibration. In *Proceedings of the SPIE Conference on Videometrics II* (Boston, MA). Vol. 2067. 21–30.

Besl, P. 1989. Range image sensors. In *Advances in Machine Vision*, J. Sanz, Ed. Springer-Verlag, New York, NY.

Besl, P. and McKay, N. 1992. A method for registration of 3D shapes. *IEEE Trans. PAMI 14*, 2 (Feb.), 239–256.

Best, L. C. 1996. Autonomous construction of three-dimensional models from range data. Ph.D. dissertation, University of Wyoming, Laramie, WY.

Bowyer, K. and Dyer, C. 1990. Aspect graphs: An introduction and survey of recent results. In *Proceedings of the SPIE Conference on Close-Range Photogrammetry Meets Machine Vision*. Vol. 1395. 200–208.

Buzinski, M. J. 1990. A comparison of touch probe and laser triangulation sensors for verifying the dimensional accuracy of manufactured parts. M.S. thesis, Purdue University, Lafayette, IN.

Connolly, C. 1985. The determination of next best views. In *Proceedings of the IEEE International Conference on Robotics and Automation*. 432–435.

Cowan, C. and Kovesi, P. 1988. Automatic sensor placement from vision task requirements. *IEEE Trans. PAMI 10*, 3 (May), 407–416.

Curless, B. and Levoy, M. 1996. Better optical triangulation through space-time analysis. In *SIGGRAPH '96*. 1–10.

Curless, B. L. 1997. New methods for surface reconstruction from range images. Ph.D. dissertation, Stanford University, Stanford, CA.

El-Hakim, S. and Beraldin, J.-A. 1994. On the integration of range and intensity data to improve vision-based three-dimensional measurements. In *Proceedings of the SPIE Conference on Videometrics III* (Boston, MA). Vol. 2350. 306–321.

Faugeras, O., Mundy, J., Ahuja, N., Dyer, C., Pentland, A., Jain, R., and Ikeuchi, K. 1992. Why aspect graphs are not (yet) practical for computer vision. *CVGIP: Image Understand. 55*, 2, 212–218.

Garcia, M., Velazquez, S., and Sappa, A. 1998a. A two-stage algorithm for planning the next view from range images. In *British Machine Vision Conference 1998*. 720–729.

Garcia, M., Velazquez, S., Sappa, A., and Basanez, L. 1998b. Autonomous sensor planning for 3D reconstruction of complex objects from range images. In *Proceedings of the IEEE International Conference on Robotics and Automation* (Leuven, Belgium). 3085–3090.

Greespan, M. 2002. Geometric probing of dense range data. *IEEE Trans. PAMI 24*, 4 (April), 495–508.

Hall-Holt, O. 1998. Technical report, Stanford University, Stanford, CA. Available online at www-graphics.stanford.edu/~olaf/nbv.html.

Hébert, P. 2001. A self-referenced hand-held range sensor. In *3rd International Conference on 3D Digital Imaging and Modeling* (Quebec City, P.Q., Canada). 5–12.

Huber, D. 2001. Automatic 3D modeling using range images obtained from unknown viewpoints. In *3rd International Conference on 3D Digital Imaging and Modeling* (Quebec City, P.Q., Canada). 153–160.

HUTCHINSON, S. AND KAK, A. 1989. Planning sensing strategies in a robot work cell with multisensor capabilities. *IEEE Trans. Robot. Automat. 5*, 6 (Dec.), 765–783.

KAHN, J., KLAWE, M., AND KLEITMAN, D. 1980. Traditional galleries require fewer watchmen. Tech. rep. RJ3021, IBM Watson Research Center, Yorktown Heights, NY.

KITAMURA, Y., SATO, H., AND TAMURA, H. 1990. An expert system for industrial machine vision. In *Proceedings of the 10th International Conference on Pattern Recognition*. 771–773.

KUTULAKOS, K. AND SEITZ, S. 2000. A theory of shape by space carving. *Intl. J. Comp. Vis. 38*, 3, 1999–218.

KUTULAKOS, K. N. 1994. Exploring three-dimensional objects by controlling the point of observation. Ph.D. dissertation, University of Wisconsin-Madison, Madison, WI.

LAMB, D., BAIRD, D., AND GREENSPAN, M. 1999. An automation system for industrial 3D laser digitizing. In *Proceedings of the 2nd International Conference on 3D Digital Imaging and Modeling* (Ottawa, Ont., Canada). 148–157.

LANGIS, C., GREENSPAN, M., AND GODIN, G. 2001. The parallel iterative closest point algorithm. In *Proceedings of the 3rd International Conference on 3D Digital Imaging and Modeling* (Quebec City, P.Q., Canada). 195–202.

MASSIOS, N. 1997. Predicting the best next view for a laser range striper system. M.S. thesis, University of Edinburgh, Edinburgh, Scottland.

MASSIOS, N. AND FISHER, R. 1998. A best next view selection algorithm incorporating a quality criterion. In *British Machine Vision Conference* (Sept. 1998). 780–789.

MAVER, J. 1995. Collecting visual information using an active sensor system. Ph.D. dissertation, University of Ljubljana, Ljubljana, Slovenia.

MAVER, J. AND BAJCSY, R. 1990. How to decide form the first view where to look next. In *Proceedings of the DARPA Image Understanding Workshop* (Pittsburgh, PA). 482–496.

MAVER, J. AND BAJCSY, R. 1993. Occlusions as a guide for planning the next view. *IEEE Trans. PAMI 17*, 5 (May), 417–433.

MILROY, M., BRADLEY, C., AND VICKERS, G. 1996. Automated laser scanning based on orthogonal cross sections. *Machine Vis. Appl. 9*, 106–118.

MOROOKA, K., ZHA, H., AND HASEGAWA, T. 1999. Computations on a spherical view space for efficient planning of viewpoints in 3D object modeling. In *Proceedings of the 2nd International Conference on 3D Digital Imaging and Modeling* (Ottawa, Ont., Canada). 138–147.

NEWMAN, T. AND JAIN, A. 1995. A survey of automated visual inspection. *Comput. Vis. Image Understand. 61*, 2 (March), 231–262.

NOVINI, A. 1988. Lighting and optics expert system for machine vision. In *Proceedings of the SPIE Conference on Optics, Illumination, and Image Sensing for Machine Vision III*. Vol. 1005. 131–136.

PAPADOPOULOS-ORFANOS, D. 1997. Numerisation geometrique automatique a l'aide d'une camera 3D de precision a profondeur de champ reduite. Ph.D. dissertation, Ecole nationale superieure des telecommunications, Paris, France.

PAPADOPOULOS-ORFANOS, D. AND SCHMITT, F. 1997. Automatic 3D digitization using a laser rangefinder with a small field of view. In *Proceedings of the International Conference on Recent Advances in 3D Digital Imaging and Modeling* (Ottawa, Ont., Canada). 60–67.

PITO, R. 1996a. Mesh integration based on comeasurements. In *Proceedings of the International Conference on Image Processing* (Lausanne, Switzerland). Vol. 2. 397–400.

PITO, R. 1996b. A sensor based solution to the next best view problem. In *Proceedings of the International Conference on Pattern Recognition*. 941–945.

PITO, R. 1997a. Automated surface acquisition using range cameras. Ph.D. dissertation, University of Pennsylvania, Philadelphia, PA.

PITO, R. 1997b. A registration aid. In *Proceedings of the International Conference on Recent Advances in 3D Digital Imaging and Modeling* (Ottawa, Ont., Canada). 85–92.

PITO, R. 1999. A solution to the next best view problem for automated surface acquisition. *IEEE Trans. PAMI 21*, 10 (Oct.), 1016–1030.

PITO, R. AND BAJCSY, R. 1995. A solution to the next-best-view problem for automated cad model acquisition of free-form objects using range cameras. In *Proceedings of the SPIE Symposium on Modeling, Simulation and Control Technologies for Manufacturing* (Philadelphia, PA.) Vol. 2596. 78–89.

PRIETO, F. 1999. Métrologie assistée par ordinateur: Apport des capteurs 3D sans contact. Ph.D. dissertation, L'Institut National des Sciences Appliquées de Lyon, Lyon, France.

PRIETO, F., REDARCE, T., BOULANGER, P., AND LEPAGE, R. 1999. CAD-based range sensor placement for optimum 3D data acquisition. In *Proceedings of the International Conference on 3D Digital Imaging and Modeling* (Ottawa, Ont., Canada). 128–137.

PRIETO, F., REDARCE, T., BOULANGER, P., AND LEPAGE, R. 2001. Tolerance control with high resolution 3D measurements. In *3rd International Conference on 3D Digital Imaging and Modeling* (Quebec City, P.Q., Canada). 339–346.

PUDNEY, C. J. 1994. Surface modeling and surface following for robots equipped with range sensors. Ph.D. dissertation, University of Western Australia, Perth, Australia.

REED, M. AND ALLEN, P. 1997. A robotic system for 3D model acquisition from multiple range images. In *Proceedings of the IEEE International Conference on Robotics* *and Automation* (Albuquerque, NM.). 2509–2514.

Reed, M. and Allen, P. 1999. 3D modeling from range imagery. *Image Vis. Comput. 17*, 1 (Feb.), 99–111.

Reed, M., Allen, P., and Stamos, I. 1997a. 3D modeling from range imagery: An incremental method with a planning component. In *Proceedings of the International Conference on Recent Advances in 3D Digital Imaging and Modeling* (Ottawa, Ont., Canada). 76–83.

Reed, M., Allen, P., and Stamos, I. 1997b. Automated model acquisition from range images with view planning. In *Proceedings of the IEEE Conference Visual Pattern Recognition* (San Juan, Puerto Rico). 72–77.

Reed, M. K. 1998. Solid model acquisition form range imagery. Ph.D. dissertation, Columbia University, New York, NY.

Roberts, D. and Marshall, A. 1997. A review of viewpoint planning. Tech. Rep. 97007, University of Wales College of Cardiff, Cardiff, U.K.

Roth, G. 2000. Building models from sensor data: an application shared by the computer vision and the computer graphics community. In *Confluence of Computer Vision and Computer Graphics*, A. Leonardis et al., Eds. Kluwer The Hague, The Netherlands.

Roui-Abidi, B. 1992. Automatic sensor placement for volumetric object characterization. Ph.D. dissertation, University of Tennessee, Knoxville, TN.

Rusinkiewicz, S. and Levoy, M. 2001. Efficient variants of the ICP algorithm. In *Proceedings of the 3rd International Conference on 3D Digital Imaging and Modeling* (Quebec City, P.Q., Canada). 145–152.

Sakane, S., Niepold, R., Sato, T., and Shirai, Y. 1992. Illumination setup planning for a hand-eye system based on an environmental model. *Advan. Robot. 6*, 4, 461–482.

Scott, W., Roth, G., and Rivest, J.-F. 2000. Performance-oriented view planning for automatic model acquisition. In *Proceedings of the 31st International Symposium on Robotics* (Montreal, P.Q., Canada). 314–319.

Scott, W., Roth, G., and Rivest, J.-F. 2001a. View planning as a set covering problem. Tech. Rep. NRC-44892. National Research Council of Canada, Institute for Information Technology, Ottawa, Ont., Canada.

Scott, W., Roth, G., and Rivest, J.-F. 2001b. View planning for multi-stage object reconstruction. In *Proceedings of the 14th International Conference on Vision Interface* (Ottawa, Ont., Canada). 64–71.

Scott, W., Roth, G., and Rivest, J.-F. 2002. Pose error effects on range sensing. In *Proceedings of the 15th International Conference on Vision Interface* (Calgary, Alta., Canada). 331–338.

Scott, W. R. 2002. Performance-oriented view planning for automated object reconstruction. Ph.D. dissertation, University of Ottawa, Ottawa, Ont., Canada.

Sedas-Gersey, S. W. 1993. Algorithms for automatic sensor placement to acquire complete and accurate information. Ph.D. dissertation, Carnegie Mellon University, Pittsburgh, PA.

Sequeira, V. 1996. Active range sensing for three-dimensional environment reconstruction. Ph.D. dissertation, IST-Technical University of Lisbon, Lisbon, Portugal.

Sequeira, V. and Gonçalves, J. 2002. 3D reality modeling: Photo-realistic 3D models of real world scenes. *Proceedings of the First International Symposium on 3D Data Processing Visualization and Transmission* (3DPVT02). 776–783.

Soucy, G., Callari, F., and Ferrie, F. 1998. Uniform and complete surface coverage with a robot-mounted laser rangefinder. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems* (Victoria, B.C., Canada). 1682–1688.

Tarabanis, K. 1992. Sensor planning and modeling for machine vision tasks. Ph.D. dissertation, Columbia University, New York, NY.

Tarabanis, K., Allen, P., and Tsai, R. 1995a. A survey of sensor planning in computer vision. *IEEE Trans. Robot. Automat. 11*, 1 (Feb.), 86–104.

Tarabanis, K., Tsai, R., and Allen, P. 1995b. The MVP sensor planning system for robotic vision tasks. *IEEE Trans. Robot. Automat. 11*, 1 (Feb.), 72–85.

Tarabanis, K., Tsai, R., and Kaul, A. 1996. Computing occlusion-free viewpoints. *IEEE Trans. PAMI 18*, 3 (March), 279–292.

Tarbox, G. and Gottschlich, S. 1995. Planning for complete sensor coverage in inspection. *Comput. Vis. Image Understand. 61*, 1 (Jan.), 84–111.

Tarbox, G. H. 1993. A volumetric approach to automated 3D inspection. Ph.D. dissertation. Rensselaer Polytechnic Institute, Troy, NY.

Tremblay, P.-J. and Ferrie, F. 2000. The skeptical explorer: A multi-hypotheses approach to visual modeling and exploration. *Auton. Robot. 8*, 2, 193–201.

Trucco, E., Umasuthan, M., Wallace, A., and Roberto, V. 1997. Model-based planning of optimal sensor placements for inspection. *IEEE Trans. Robot. Automat. 13*, 2 (April), 182–194.

Urrutia, J. 2000. Art gallery and illumination problems. In *Handbook of Computational Geometry*, J.-R. Sack and J. Urrutia, Eds. Elsevier, New York, NY.

Whaite, P. and Ferrie, F. 1990. From uncertainty to visual exploration. In *Proceedings of the 3rd International Conference on Computer Vision* (Osaka, Japan). 690–697.

Whaite, P. and Ferrie, F. 1991. From uncertainty to visual exploration. *IEEE Trans. PAMI 13*, 10 (Oct.), 1038–1049.

WHAITE, P. AND FERRIE, F. 1992. Uncertain views. In *Proceedings of the IEEE Conference on Visual Pattern Recognition.* 3–9.

WHAITE, P. AND FERRIE, F. 1997. Autonomous exploration: Driven by uncertainty. *IEEE Trans. PAMI 19*, 3 (March), 193–205.

WHEELER, M. D. 1996. Automatic modeling and localization for object recognition. Ph.D. dissertation, Carnegie Mellon University, Pittsburgh, PA.

WIXSON, L. E. 1994. Gaze selection for visual search. Ph.D. thesis, University of Rochester, Rochester, NY.

XIE, S., CALVERT, T., AND BHATTACHARYA, B. 1986. Planning views for the incremental construction of model bodies. In *Proceedings of the 8th International Conference on Pattern Recognition* (Paris, France). 154–157.

YE, Y. 1997. Sensor planning for object search. Ph.D. dissertation, University of Toronto, Toronto, Ont., Canada.

YE, Y. AND TSOTSOS, J. 1999. Sensor planning for 3D object search. *Comput. Vis. Image Understand. 73*, 2 (Feb.), 145–168.

YI, S. 1990. Illumination control expert for machine vision: A goal driven approach. Ph.D. dissertation, University of Washington, Seattle, WA.

YI, S., HARALICK, R., AND SHAPIRO, L. 1995. Optimal sensor and light source positioning for machine vision. *Comput. Vis. Image Understand. 61*, 1 (Jan.), 122–137.

YU, Y. AND GUPTA, K. 2000. An information theoretic approach to view point planning for motion planning of eye-in-hand systems. In *Proceedings of the 31st International Symposium on Robotics*. 306–311.

YUAN, X. 1993. 3D reconstruction as an automatic modeling system. Ph.D. dissertation, University of Alberta, Edmonton, Alta., Canada.

YUAN, X. 1995. A mechanism of automatic 3D object modeling. *IEEE Trans. PAMI 17*, 3 (March), 307–311.

ZHA, H., MOROOKA, K., AND HASEGAWA, T. 1998. Next best viewpoint (NBV) planning for active object modeling based on a learning-by-showing approach. In *Proceedings of the 3rd Asian Conference on Computer Vision* (Hong Kong). 185–192.

ZHA, H., MOROOKA, K., HASEGAWA, T., AND NAGATA, T. 1997. Active modeling of 3D objects: Planning on the next best pose (NBP) for acquiring range images. In *Proceedings of the International Conference on Recent Advances in 3D Digital Imaging and Modeling* (Ottawa, Ont., Canada). 68–75.

Received March 2002; revised September 2002; accepted November 2002