Summer 6-1999

A Smart Move: *A System To Intelligently Model and Control Robots in Extreme Environments*

Surya Priyadarshi Singh
University of Tennessee - Knoxville

A SMART MOVE

A System To Intelligently Model and Control Robots in Extreme Environments

Surya Singh

March 19, 2000

Abstract

The extensive use of robotics in hazardous environments has been limited by the inability to easily model and control robotic systems with high accuracy. In general, purely manual robot modeling systems lack the precession needed and fully autonomous systems are unable to cope with high levels of environmental variability or abnormalities. It is the objective of this research effort to develop and use a computational modeling and control system consisting of a series of sensors and algorithms that will augment manual operations to not only minimize error, but to reduce the tedium and difficulty presently characteristic of this form of robotic operation.

The basic methodology of this system is to extend the traditional master-slave telerobotic model via a series of sensors coupled to a computational model of the environment, which together adjust the velocity scaling of the end-defector. In particular, this system is implemented using a set of transformations that adjust the velocity scaling of the end-defector in proportion to the distance from the end defector to the modeled object or location.

To design and test this system an experimental setup consisting of a multi-component mockup wall, a Titan robot arm, and a mini-master joint space controller. The experiment measured the time it took to select four rectangular arranged points on the surface of this wall. System performance was evaluated on the time necessary to accurately position and navigate the end-defector around the four points and the number of collisions and errors occurring throughout the process.

The results of these experiments show that the algorithm tested resulted in a significant increase in task execution efficiency as revealed by the reduction in mean execution time, frequency/magnitude of collisions, and reduction in operator fatigue. Thus, the experiment shows that an assisted robotic modeling and control strategy maybe a viable method for improving operator interactions with the robot system.

Acknowledgments

This research was performed primarily at the University of Tennessee's REmSL (Robotics and Electromechanical Systems Laboratory). The research described herein was assisted by the faculty and students associated with this laboratory. Some of the many individuals who have contributed to this research effort include Dr. Everett, Dr. Hamel, Dr. Widner, Dr. Thayer, Mr. S. Hale, Mr. S. Mitchell, and Mr. X. Fu.

This work was performed under contract DE-AR26-97FT34314 for the Department of Energy, Federal Energy Technology Center, Morgantown, West Virginia. The University of Tennessee and Carnegie Mellon University teams acknowledge and appreciate the interest, guidance, and constructive support of the DOE contracting officer's representative, Mr. Vijendra P. Kothari. His professionalism and kindness have made all the difference to us. The research teams further acknowledge for the continuing technical support of Dr. Linton Yarbrough and Mr. John DeGregory of the DOE Robotics Technology Development Program and Dr. Paul Hart of the DOE Deactivation and Decommissioning Focus Area. The endorsement that these men have for the technical vision of the project is essential. Finally, Mr. Dennis Haley and Mr. Robert Barry of the Oak Ridge National Laboratory have provided technical input and assistance in numerous ways especially in arrangements for loan equipment.

Introduction

It has been long established that the productivity of remote handling is a key problem in nuclear operations. One way to enhance remote productivity is to incorporate integral task automation such that specific tasks can be accomplished faster and with higher work quality than standard teleoperations. This productivity issue is the underlying motivation for the Robot Task Space Analyzer (RTSA).

The RTSA is an enabling technology necessary to allow mobile remote work systems to utilize automated operations in unstructured task environments which are typical in the environmental remediation challenges associated with the US Department of Energy's nuclear facilities. The RTSA is a human interactive system which allows a remote operator to direct the construction of 3D geometrical descriptions of task objects (e.g., pipes, valves, tanks, etc.). In its present form, RTSA uses stereo remote vision and laser range sensors to gather physical data about objects. Computer algorithms directed by the operator then compute the locations and orientations of the objects for subsequent use in automated robot task planning and execution.

As piping components (e.g., pipes, elbows, flanges, and tees) as well as process equipment (e.g., tanks and valves) have standardized sizes, the RTSA system incorporates pre-constructed models of piping components. The pre-constructed models enhance autonomous scene analysis and reduce manual modeling time. Upon the operator's request, a computer algorithm using information obtained from the stereo camera images or the laser range camera data computes the locations and orientations of specific components specified by the operator. While the operator continues manual task space modeling, the autonomous algorithms attempt to find scene objects as background processes. The most promising aspect of the RTSA system is that it builds an in situ model of unstructured task environments that are typical of environmental remediation associated with the U.S. Department of Energy's nuclear facilities. The use of RTSA technologies may allow greater operator productivity by reducing fatigue. For example, one of the principal causes of operator fatigue – the repetitive fine control of remote manipulator teleoperation tasks – is addressed and mediated by this system. The RTSA system also uses modern and intuitive graphical user interfaces to aid the operator's model building task.

The RTSA project revealed the benefits of Human Interactive Stereo (HIS) and Semi-Autonomous Range (SAR) scene analyzes can be practically combined to achieve the following goals:

- Extensions of SAR to include direct model manipulation, processing of cluttered environments, and simultaneous recognition and processing of multiple objects
- Extensions of HIS to include background semi-autonomous operation
- Human Machine Interface improvements which include streamlining the operator's GUI, implementation of human factors research, and inclusion of a model library of standard process objects

The RTSA project indicates that it is feasible and practical to construct accurate geometrical models quickly and easily. Further research will employ performance metrics to quantify the speed and accuracy of this modeling and control approach.

Contents

List of Figures

List of Tables

List of Acronyms

2D	Two Dimensional
3D	Three Dimensional
BNC	Bayonet Connector
CCD	Charge-Coupled Device
CMU	Carnegie Mellon University
DOE	Department of Energy
ER&WM	Environmental Restoration and Waste Management
FORM	Free-form Object Recognition Method
GUI	Graphical User Interface
HIS	Human Interactive Stereo
ICP	Iterative Closest Point
IGRIP	Interactive Graphical Robotic Instruction Program
LASAR	Laser Range Camera
LoG	Laplacian of Guassian
MHz	Megahertz
MIL	Matrox® Imaging Library
OD	Outer Diameter
OOI	Object of Interest
ORNL	Oak Ridge National Laboratory
PCI	Peripheral Component Interconnect
PPB	PCI to PCI Bridge
PV	Panoramic View
QPSM	Quadric/Planar Segmentation and Matching
R&D	Research and Development
RGB	Red Green Blue
ROI	Region of Interest
RTSA	Robot Task Space Analyzer
SAR	Semi-Autonomous Range analysis system
SGI	Silicon Graphics®, Inc.
TSSA	Task space scene analysis
VME	Versabus Module European

Nomenclature

Roman

a	Horizontal distance between camera tilt axis and the focal array
a_i	Discretized intervals of variation, $i \in [1, p]$
A	Midpoint of two focal points of the stereo camera array
$A(s_i)$	Array of integer accumulators, $i \in [1, p]$
A_l	Focal point of the left camera
A_r	Focal point of the right camera
b	One half of the baseline camera separation (baseline $= 2b$)
c	Vertical distance between origins of the camera and world frames
C	Similarity function for spin image comparison
Ci_j	j^{th} object shape in i^{th} image, $i \in [1, 2]$, $j \in [a, p]$, where $a \equiv$ actual and $p \equiv$ predicted
d	Vertical distance from camera focal points to origin of camera frame
d_{gc}	Normalized distance between spin map coordinates
D_{gc}	Geometric consistency measure
$E(u, v)$	Edge image
f	Camera focal length
$f(x, a)$	Target shape parameterization
L	List of likely model scene point correspondences
N	Number of overlapping bins
P	Scene point in the workspace
P'	Auxiliary point used in locating scene point P. (See Figure **??**)
${}^{i-1}P_i$	3×1 Position vector of origin O_i relative to origin O_{i-1}
P_l	Projection of scene point P onto the focal array of the left camera
P_r	Projection of scene point P onto the focal array of the right camera
R_p	Rotation of a pipe about its center
$\vec{R}$	Vector from camera coordinate frame A to scene point P
${}^{i-1}R_i$	3×3 rotational transformation from frame Σ_{i-1} to frame Σ_i
s_i	Sizes of the discretized intervals of variation, $i \in [1, p]$
S_t	Stereo pose estimation minimization parameter
T_{gc}	Correspondence grouping threshold
${}^{i-1}T_i$	Homogeneous 4×4 tranformation from frame Σ_{i-1} to frame Σ_i
(x'_l, y'_l)	Scene point expressed in left camera image coordinates
(x'_r, y'_r)	Scene point expressed in left camera image coordinates
(x_l, y_l, z_l)	Scene point expressed in left camera coordinates
(x_r, y_r, z_r)	Scene point expressed in right camera coordinates
W_{gc}	Weighted measure of geometric consistency
(X_p, Y_p, Z_p)	Coordinates of pipe center
${}^{i-1}X_i$	Unit vector in the direction of X_i, expressed in frame Σ_{i-1}
${}^{i-1}Y_i$	Unit vector in the direction of Y_i, expressed in frame Σ_{i-1}
${}^{i-1}Z_i$	Unit vector in the direction of Z_i, expressed in frame Σ_{i-1}

Greek

$c\theta_p$	Shorthand notation for $\cos(\theta_p)$
$c\theta_t$	Shorthand notation for $\cos(\theta_t)$
$s\theta_p$	Shorthand notation for $\sin(\theta_p)$
$s\theta_t$	Shorthand notation for $\sin(\theta_t)$
λ	Variance weighting variable
θ_p	pan rotation angle
θ_t	tilt rotation angle
Σ_i	Frame "i", where $i \in [1, 4]$
τ	User defined threshold

1 General Background and Review of Literature

The project deals with the use and effects of computer-assisted robotics algorithms to facilitate robotic control in hazardous or life-threatening environments. In particular, this experiment studies the use of a computer-assistance algorithm coupled with virtual modeling tools to aid teleoperation by reducing execution time, errors, and fatigue.

Telerobotics is the execution of robotics operations and tasks over some distance. Telerobotics extends a person's sensing and control capabilities to any remote location or area, which otherwise may be inaccessible. The application of telerobotics ranges from a mechanical gripper arms to the electronic and highly automated systems that extend the user anywhere on earth or space (e.g., Mars Pathfinder).

Telerobotic systems, originally developed in the late 1950's, consist of a series of separate mechanical input and output devices linked together. Modern-day telerobotics systems, while high advanced, share a number of features and characteristics with traditional systems, such as:

- Telerobotics generally involve systems where a remote manipulator is controlled by a human operator in a linear, direct, and continuous fashion.
- The operator's controls, which may provide force feedback and other sensory extensions, are referred to as the Master; the robotics systems which performs the actual manipulation is referred to as the Slave.
- Computers have traditionally been used as an intermediary between to the Master inputs and the Slave outputs.

In general, manual telerobotic systems lack the precision needed while fully autonomous systems are unable to cope with high levels of environmental variability or abnormalities. It is the objective of this research effort to use a computational model and set of remote sensors to augment manual telerobotic operations to not only minimize error, but to reduce the tedium and difficulty characteristic of present day precise telerobotic operations.

Due to the difficulties inherent in both purely manual and purely autonomous control systems, a great deal of interest has been placed in strategies that merge human decisions with computer assistance. One of the first techniques developed involved averaging velocity commands from the Master with those from an automatic controller. Another approach tried involved using sensors and computer models to maneuver the manipulator away from obstructions automatically regardless of the input parameters.

Two recent development in assistance algorithms are Combined Automation and the Variable Velocity Scaling (VVS) algorithm. Combined Automation is the automated use of multiple simultaneous modeling approaches whose oversight and selection is directed by the operator. This technique is main thrust of the RTSA system as it increases accuracy and reduces fatigue by not having the operator repeat mundane tasks. This algorithm monitors the position of the Slave manipulator and uses this information in conjunction with a computation model of the environment to continuously vary the velocity scaling between the Master controller and the robot manipulator. This algorithm gives the user more precise control of the manipulator as it approaches the object of interest. The VVS algorithm allows the user to follow the contour of a geometry and it assists (but does not necessarily prevent) the user in avoiding collisions. In addition, the VVS system will only enhance, rather than supersede, the human operator's inputs. Moreover, this algorithm should significantly reduce operator fatigue by reducing the turnaround time, tedium, and errors often associated with purely manual telerobotics.

A crucial component of most autonomous and computer-aided telerobotic systems is the presence of a precise mathematical/computer model of the remote environment. This quantitative data about the systems are needed by the assistance algorithms to fully support the user in the completion of the tasks. While many three dimensional camera and other related sensing systems have been developed, none of these can produce a accurate mathematical model that can be readily used by this algorithm. However, many of the geometries present in remote or high precision environments can be modeled as a series of standard geometries.

This research will advance the variable velocity algorithm by showing that it can be successfully applied to a series of basic models and conversely to a variety of telerobotic applications ranging from environmental restoration to highly precise operating theaters with little margin for error.

2 Technical Background and Overview

Many environmental restoration and waste management (ER&WM) challenges involve radiation or other hazards which will necessitate the use of remote operations to protect human workers from dangerous exposures. The mandate to work remotely carries the implication of substantially greater costs to complete this work agenda since remote work systems are inherently far less productive than direct human worker equivalents due to the inefficiencies of teleoperation. To reduce costs and improve quality, much attention has been focused on methods to improve the productivity of combined human operator/remote equipment systems, and the achievements to date are modest at best. Since the ER&WM agenda can not be postponed, the most promising avenue in the near term is to supplement conventional remote work systems with robotic planning and control techniques borrowed from manufacturing and other domains where the impact of automation has been lower cost and improved quality of workmanship. Such a combination of teleoperation and robotic control will yield telerobotic work systems that outperform currently available remote equipment.

It is important to note that the basic hardware and software features of most modern remote manipulation systems can readily accommodate the functionality required for telerobotics. Further, several of the additional system ingredients necessary to implement telerobotic control — machine vision, 3D object and workspace modeling, automatic tool path generation and collision-free trajectory planning — are existent.

Practical and reliable implementation of telerobotic systems in ER&WM contexts is an unrealized objective, despite the potential payoff of telerobotics. This can be attributed to several formidable technical challenges unique to field automation. Almost always the geometry of the task environment is highly unstructured and uncertain. Likewise, the precision and accuracy of the requisite geometric knowledge varies from task to task, as does the extent of the task space itself.

Finally, a significant fraction of the tasks to be performed are complex by any standard. These factors put full automation of ER&WM tasks beyond the reach of current technology. However, there are certain subtasks that are amenable to automatic planning and execution; interjection of telerobotic subtasks into the overall sequence is the most attainable exploitation of automation's benefits in the foreseeable future. Near term implementation of telerobotic capability in a typical ER&WM application will thus manifest itself in an operational sequence such as that as depicted in Figure 2-1.

Automation of a task requires complete quantitative data about the task/subtasks to be performed, the manipulation systems, and the tooling systems to be used. Task space scene analysis (TSSA) refers to the process by which the remote work system gathers geometrical and other types of information that are necessary to characterize, analyze, and plan the automated task execution. For example, in a dismantlement scenario the task may be to remove a segment of process piping using remote manipulators and cutting tools. If such a task is to be automated, it is necessary to describe the location and orientation of each piping element with respect to the remote work system. This data representation, or model, must be complete and accurate to an extent dictated by the specific tool being used: positioning of a shear demands less accuracy than achieving the proper standoff for a plasma arc torch. Once a sufficient model is available, planning the manipulator motions is straightforward and the cutting can be automatically executed with confidence and reliability.

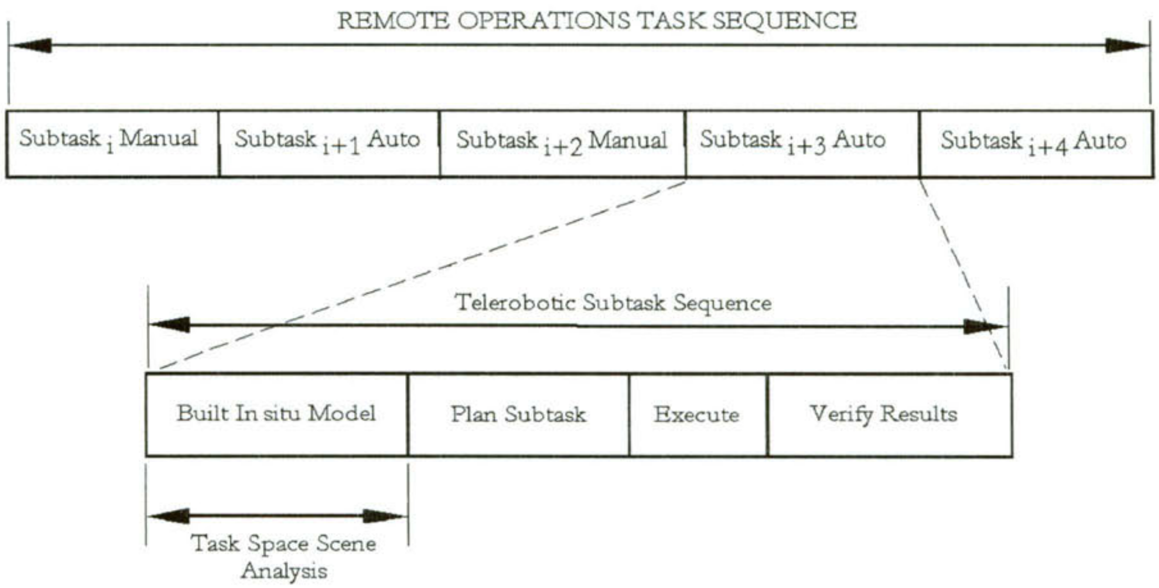

Figure 2-1: Telerobotics Operations Cycle

The Robot Task Space Analyzer (RTSA) is a system which performs TSSA, and is in essence a model builder of the near-field of view of the mobile work system. Unlike the notion of world model building, RTSA functions in the region of "space" in the near-field which is within the sphere of influence of the remote work system where the current task operations are to be performed. RTSA performs an integral step in the telerobotics operations cycle and it must exhibit a level of efficiency that allows telerobotic execution to provide performance benefits over conventional teleoperational execution.

As depicted in Figure 2-1, telerobotic execution requires a "programming" phase and an "execution" phase for each task to be performed. The programming phase is the RTSA function plus task planning; it is the most important part of the operation since subsequent execution is fully automatic and can progress at the maximum speed of the remote hardware. Techniques to automatically plan optimal tooling paths are well established through decades of robotics research in the manufacturing domain. Therefore, RTSA is an enabling technology which determines the ultimate overall performance of any telerobotics concept.

The RTSA project approach emphasizes testing and evaluation of results to steer the evolution of component technologies and the system as a whole. Several scenarios have been considered for testing and evaluation when the integrated RTSA system will be tested with a actual remote work system. The set of scenarios have been chosen to be representative of typical decontamination and dismantling operations and to span the set of task space conditions (e.g., lighting, surface conditions, etc.) expected in realistic situations. The scenarios which have been included are:

- object manipulation, such as pipe cutting, waste segregation, or retrieval of a partially buried waste container, either with clear access to the object or with clutter in front of or around it

- decontamination of surfaces, including large, accessible patches such as vessels and tank walls, and small surface areas in cluttered environments such as hot cells and glove boxes.

Laboratory scale tests will be conducted under controlled conditions to measure the RTSA model building efficiency for these scenarios. It should be noted that these objectives also encompass task space scene analysis requirements that are associated with tank wastes, mixed waste operations, and many other areas connected with ER&WM.

The technical areas that have been investigated were identified based on an analysis of the current SAR system and in light of the task scenarios described above. Specific objectives were:

- Improve the present Artisan system's ability to recognize individual objects. Considerable improvements have been made by building upon the fundamental algorithms developed previously. Most importantly, steps have been taken to increase the precision of object pose and dimension estimates made by SAR, and the number of false matches has been reduced.

- Make scene analysis applicable to tasks and scenarios where there are large numbers of objects of various types. Extensions of the Artisan's underlying algorithm have been made to allow simultaneous recognition of multiple objects and merging of several range images into a single 3D data set.

HIS has been extended in three areas to increase its utility and performance: automation of the scene decomposition process, application to laser range images directly, and background search and recognition of indicated objects of interest.

- *Automate the scene decomposition process.* Earlier work showed that a major operator workload resulted from having to reposition the narrow field of view stereo camera pair for the next scene patch. The objective of the effort has been to develop automatic controls that allow the operator to specify a large region for analysis while the HIS controls system helps subdivide the image for specific objects (e.g., valves, elbows, pipes, etc.) of interest.

- *Extend to laser range images.* Typical laser range cameras provide both 3D range and intensity images. It is clear that an operator can interact with these images in a manner equivalent to HIS. The objective of this effort has been to extend the HIS tool kit to allow the operator to graphically interact with the laser range images to construct task models directly.

- *Background search and recognition of indicated objects of interest.* The performance of computer vision object recognition algorithms can be dramatically influenced by a priori knowledge and search space foci. The objective of the activity has been to incorporate automated schemes that operate in parallel and in the background with HIS to search for and identify the objects that the operator has specified.

Our intent has been to simplify operator access to the RTSA functions described above, hence the human-machine interface design has been a major focus of attention and innovation. The RTSA is designed to allow the human operator to direct and participate in the task scene analysis. In general, this involves reducing the level of expertise required to properly use the scene analysis techniques and alleviating the operator's administrative workload.

To realize a truly useful integrated task space scene analysis tool, it is necessary to objectively evaluate the component technologies of which RTSA is comprised. This will include making quantitative assessments for all of the model building modules that are incorporated. Metrics that will be evaluated are:

- the percentages of time modules correctly recognize objects
- the percentages of time modules fail to recognize or report wrong matches
- the accuracy with which modules report dimensions, location and orientation of recognized objects
- how fast the modules work
- what pathological scene conditions result in failures

Beyond this raw assessment of performance, test results are being used to help apply RTSA as a remote operator's tool. In particular, it shows which method works best given scene conditions and the set of objects to recognize and how to optimally set parameters associated with each algorithm.

The success criteria corresponding to the RTSA project technical objectives are summarized in Table 2-1. These criteria will be used to evaluate the results obtained in the actual work.

Table 2-1: Success Criteria

	Criterion
Extensions of Semi-Automatic Range Image Analysis	Automatic selection of mesh resolution
	merging of simple data sets
	automatic setting of algorithm parameter values
	simultaneous recognition of multiple objects
Extensions to HIS	semi-automatic operation
	human-interactive task-space modeling from laser range images
	background HIS for automatic object modeling
Human-Machine Interfaces	user choice of interaction with video or range images
	GUI for simultaneous recognition of multiple objects
	GUI for mesh merging
	common library of object models
	video overlays on 3D objects and projection of objects onto 2D images
Laboratory Testing	statistics on RTSA accuracy
	statistics on RTSA speed
	accuracy/speed tradeoffs quantified
Integration	software module functions & I/O documented
	system requirements documented
	unified user interface
	interfaces to several range sensors
	ready to begin integrated testing
Integrated testing	comparative data on task performance with and without RTSA
	evaluation by remote equipment operators
	estimates of the potential cost savings

3 Problem

To develop and evaluate a computer-assisted robotic control system that will control and position robotic systems with high accuracy for use in hazardous and/or life threatening environments.

4 Purpose

The purpose of this project is to use a computational model and set of remote sensors to augment manual telerobotic operations to not only minimize error, but to reduce the tedium and difficulty which presently characterizes precise telerobotic operations.

5 Hypotheses

It is postulated that the use of a computer assisted robot control strategy where the operator remains in full control should result in rapid and high precision telerobotics. This system should not only yield more accurate control, but will do so more rapidly and with less user fatigue. This project experimentally investigates this system in terms of mean execution time and number of collisions.

6 Research Protocols and Procedure

The RTSA Project has produced in situ geometrical model building functionalities based on semi-autonomous range image analysis, human interactive stereo, and enhanced human machine interfaces. The basic technical approaches followed in the development of each of these functionalities are summarized below, including plans for laboratory testing and evaluation of the results. Section 7 contains a detailed description of the RTSA design results.

6.1 Extensions of Semi-Autonomous Range Image Analysis

Starting with the existing technology developed for Artisan, experiments were conducted to evaluate the system's capabilities and develop additional enhancements based on those results. Testing was conducted through controlled trials in the laboratory. The evaluation was based on a set of metrics including quantitative characterization of the accuracy of object recognition and registration, as well as qualitative evaluation of the adequacy of the approach in the context of remote operations.

Improvements in Recognizing Individual Objects

Two different object recognition algorithms have been developed for Artisan. The first method (Quadric/Planar Segmentation and Matching, or"QPSM") segments the 3D surfaces into planar and quadric patches and matches the resulting scene description to analogous descriptions of object models in a database (developed off-line from CAD descriptions of objects). The other method (Free-form Object Recognition Method or "FORM") is based on a technique known as geometric indexing. In this case a collection of 3D surface points is transformed into a set of 2D representations, called spin images, that describe the spatial relationship of each point to all the others. The stack of spin images representing the scene data are then compared to stacks of spin images of models in the database to arrive at a few number of plausible correspondences. Each of these is further refined using a modified iterative closest point (ICP) algorithm that outputs the optimal estimate of the recognized object's dimensions, location and orientation in the task space.

The internal representation used in the FORM can be statistically evaluated in order to compute the best parameters for object recognition. A histogram of correlation between the spin maps of a given model is currently used to determine the saliency of particular points of the model, and to help compute thresholds for recognition. A similar technique can be used to compute the saliency of the spin maps for various densities of surface points and to compute the optimal data resolution for a particular object. Such an algorithm has been developed for use off-line during model generation.

The quality of index-based registration is dependent on the number of points used on model and scene data sets, i.e., the resolution of the 3D meshes used in the recognition, and the appropriate resolution varies depending upon the object. The resolution is currently set manually though it can be computed automatically from the models. Furthermore, the current approach uses a reduced data set in the final registration step. While increased data density improves registration accuracy, there is a tradeoff between the improved accuracy and the increased execution time due to the larger number of points used. Therefore, for full resolution data sets the gain accuracy were obtained and strategies were developed for choosing the appropriate data density based on the tradeoff analysis.

Parameterization of object models is useful for constructing a task space scene model. For example, the scale factors along arbitrary axes can be recovered by the recognition algorithm and adjusted by the operator to stretch or shrink the object. However, the indexing approach does not allow such parameterization and needs to be extended to do so. This can be achieved by developing a means to measure the scale of an object, and scaling spin images appropriately. Alternatively, a method to compare spin images of different scale can be developed. Both approaches will be implemented and compared in terms of computational efficiency.

The two Artisan object recognition approaches perform differently on different classes of objects. For example, QPSM performs better on simple objects with full symmetries (e.g., pipes) and on objects that are almost-polyhedral, while FORM is better suited to recognize complex shapes like valves. Therefore, strategies for switching between recognition modes are needed for optimal use of Artisan. The main criteria for deciding among recognition strategies are symmetries, complexity of the object (as measured by the errors in the quadric and planar segmentations), and density of points required in FORM. Those criteria can be evaluated from the model mesh and combined into one or several metrics used for deciding among recognition modes.

6.1.1 Extension to Simultaneously Recognize Multiple Objects

The current Artisan requires the user to run the recognition algorithm for each model separately. An obvious improvement at the system level is to make it possible to recognize several objects in a single step. The user should be able to specify large sets of object models and to let the recognition algorithm determine which objects are present in the scene. Further, Artisan should be able to process several recognition operations concurrently in order to relieve the operator from the burden of searching through the database and selecting objects one by one.

The geometric indexing approach was designed with simultaneous multi-object matching in mind. Multiple indexing tables (corresponding to multiple models being tested in a given run) can be grouped into a single table. This combined table can be used during the recognition step as if it were a single model. After indexing, the best sets of point matches can be labeled as different objects to allow simultaneous identification of multiple objects in a scene. Encouraging preliminary results on multi-object matching have been obtained off-line using complex objects from the medical domain.

Those techniques have been expanded upon to realize simultaneous multi-object recognition. This task addresses fundamental issues such as the structure of the combined table and the control algorithm for dividing sets of point matches into subsets corresponding to different objects. It also addresses operational issues such as controlling the size of the combined table and managing multiple recognition thresholds. An operational capability for simultaneous multi-object recognition results, and this in turn reduces the operator's effort required to construct the task space model.

6.2 Extensions to Human Interactive Stereo

The accuracy of task space models generated by stereo-ranging is determined by a number of factors including camera calibrations, system pointing accuracies, etc. To achieve maximum accuracy, it is necessary to focus the cameras on a relatively narrow field of view of about 2 feet square. In most tasks, the region of interest is much larger than this narrow window making it necessary to patch together multiple sub-regions of interest within the HIS window. New tools to automate camera field of view selection and combine multiple objects of interest have been developed which allow the operator to specify objects of interest within a large region of interest while the system automatically zooms in on each designated object of interest.

6.2.1 Extension of HIS with Range Images

The HIS modeling technique can be extended to graphical overlay processes with laser range camera images in addition to stereo camera views. The HIS GUI and graphics tools will be adapted to be used with range images. The operator will specify object tie points in the range image which will be used in conjunction with the range data set, 3D analysis, and the object specifications to define and locate the object model. The overall model building process will follow the scheme used in HIS.

6.2.2 Background HIS

The basic approach associated with background HIS is to allow the computer to attempt to model well defined task space components in parallel with other operator actions. If the computer algorithms are successful, then time has been gained. If they are unsuccessful, the remaining objects will be handled manually by the operator. A time limit (on the order of 5 minutes) will be placed on background HIS processing with the premise that if the computer vision efforts are likely to be successful, they will obtain results fairly quickly.

6.3 Development of Human Machine Interfaces

New operator interfaces have been developed in three areas: operation of SAR, operation of HIS, and operation of RTSA at the system level.

Artisan and HIS each presently have independent user interfaces designed to support their idiosyncrasies. An integrated system, however, requires seamless combination of elements of both, as well as new functions, in a unified interface. The windows within which the user controls sensors, modifies algorithm parameter values, views images, and sees results should have a consistent look and feel.

Additional work is needed to make better use of Envision® as a geometry, reasoning and planning engine. Envision® is the Deneb® 3D kinematic modeling package. For instance, a custom Envision® menu to retrieve models of the remote work system, the tools to be used, and motion scripts for the task needs to be developed. Envision® provides the development environment for custom menu creation and the hooks for linking external software processes; these methods have been used in both Artisan and HIS and will be expanded upon. Once the user inputs this information, it can be used to guide the TSSA process. As scene analysis proceeds, the user should be able to overlay video onto the Envision® display and vice versa. This will improve his ability to assess correctness of the resulting scene model. Mapping of video onto the Envision® work cell is a capability currently under development at Deneb®; the basics of projecting 3D object models onto 2D video and range image displays exist in the current HIS and SAR systems, respectively, but need to be refined.

6.4 Experimental Testing

The first battery of laboratory tests will focus on raw recognition performance using small groups of objects. Ground truth will be established by making accurate measurements of all object dimensions and their location relative to the sensor. To the maximum extent possible and practicable, surveying equipment and/or optical benches will be used to acquire ground truth measurements.

Results reported by the TSSA techniques will be compared to ground truth to generate statistics of measurement accuracy. During these tests, parameters associated with the TSSA algorithms will be varied to generate accuracy versus parameter value statistics. A second battery of tests will use larger collections of objects and will focus on systemic performance such as execution time and the benefits of range data merging. Following the tests, statistical analysis will be used to evaluate optimum parameters, tradeoffs between speed and accuracy, and which technique is most useful given a particular type of object in the task space scene.

Tests using full scale mock-ups will be conducted to measure systemic performance, including execution time, the effects of object count and the effects of occlusions within the scene. Statistical analysis will be used to evaluate tradeoffs between speed and accuracy and to identify the most appropriate technique given a particular type of object in the task space scene.

Successful completion of this task will be evident when:

- Statistics on measurement accuracies for a large number (≥ 50) objects of varying size and shape according to the RTSA technique applied have been generated.
- Statistics on time required for construction of a task space model according to the type of object and RTSA technique have been generated.
- Tradeoffs between accuracy and time required using each technique have been identified.
- Optimal values of parameters associated with each algorithm have been established and methods to set them automatically have been created.

7 Research Results and Analysis

The principal technical work of the RTSA Project has been completed and is described in this section. The overall functional architecture of the RTSA is presented in Section 7.1. This architecture encompasses features necessary to allow effective real-time manual model construction with automated background processing of objects of interest from either stereo or range images. Sections 7.2 and 7.3 provide detailed design descriptions of the human interactive stereo and semi-autonomous range analysis systems. The prototype RTSA hardware configuration is summarized in Section 7.4. Laboratory testing utilizing a realistic ER&WM process equipment mockup will be performed in the near future. The test program is presented.

7.1 RTSA Functional Architecture

The arrangement of the overall RTSA functional architecture, as depicted in Figure 7-1, is designed such that the entire Panoramic View to be modeled may be attacked by several different modules simultaneously while minimizing the burden on the operator.

Autonomous object recognition is used in parallel with a manual operation mode such that overall modeling cycle time is minimized. The inherent trade-off with autonomous scanning involves accuracy versus computational complexity and processor burden; the RTSA functional architecture allows use of human interaction for assistance with the more complex regions in such a way that the autonomous modules may be used mainly in regions containing objects that have a high rate-of-success for auto recognition. The RTSA takes advantage of the operator intuition and experience and utilizes it such that modeling accuracy and speed are increased and operator burden is decreased. The overall effectiveness of this functionality can be iterated during testing to continually improve the effectiveness of the overall system in terms of both accuracy and usability.

The RTSA functional architecture has been designed to minimize the amount of time necessary to construct the geometrical model of a given task scene by incorporating a combination of operator-controlled modeling and automated modeling. The best way to describe architecture is to walk through the basic steps involved in developing a task space model. This step began by assuming that the remote work system has been staged at the work site and that the RTSA imaging sensors are trained on a panoramic view (PV) of an overall region to be modeled. At this point, the operator has a live video image of the PV on his RTSA terminal that is subsequently captured with the frame grabber. This captured PV image is used throughout the modeling process for the purpose of bookkeeping the placement of the individual component models.

To initiate the modeling process, the user selects multiple regions of interest (ROI) by enclosing the individual regions with the computer pointing device. Each ROI will invariably contain one or more of several standard process components (pipes, elbows, tees, flanges, etc.) that the operator recognizes and wishes to model. The complexity of the ROI will dictate a modeling method to be used for the region as specified by the operator. The decision as to whether the region should be scanned autonomously (using stereo or range information) or scanned manually (via stereo data) is left up to the operator and is based upon his intuition and experience. It is envisioned that user intuition for a priori prediction of the effectiveness of the AutoScan algorithms operating on regions of varying complexity will grow rapidly with continued testing of the RTSA system. Incorporation of human interaction is a key factor for allowing the RTSA system to generate geometrical component models in a time-efficient manner. Also paramount to the effectiveness of the RTSA is the ability to work on multiple ROI within the PV simultaneously. The RTSA workload is partitioned between autonomous scanning modules operating in the background and a manual module that is continuously available in the foreground for use by the operator. In this manner, the overall PV model is attacked from several vantage points at once; operational redundancy of this form will undoubtedly expedite the completion of the overall geometric model. Modeling results from both the manual and AutoScan processes are placed in the PV image as they are completed, refined, and approved by the operator. It must be emphasized that the final decision concerning the acceptance (i.e. placement) of component models - regardless of whether the models were generated autonomously or manually - is left solely to the operator.

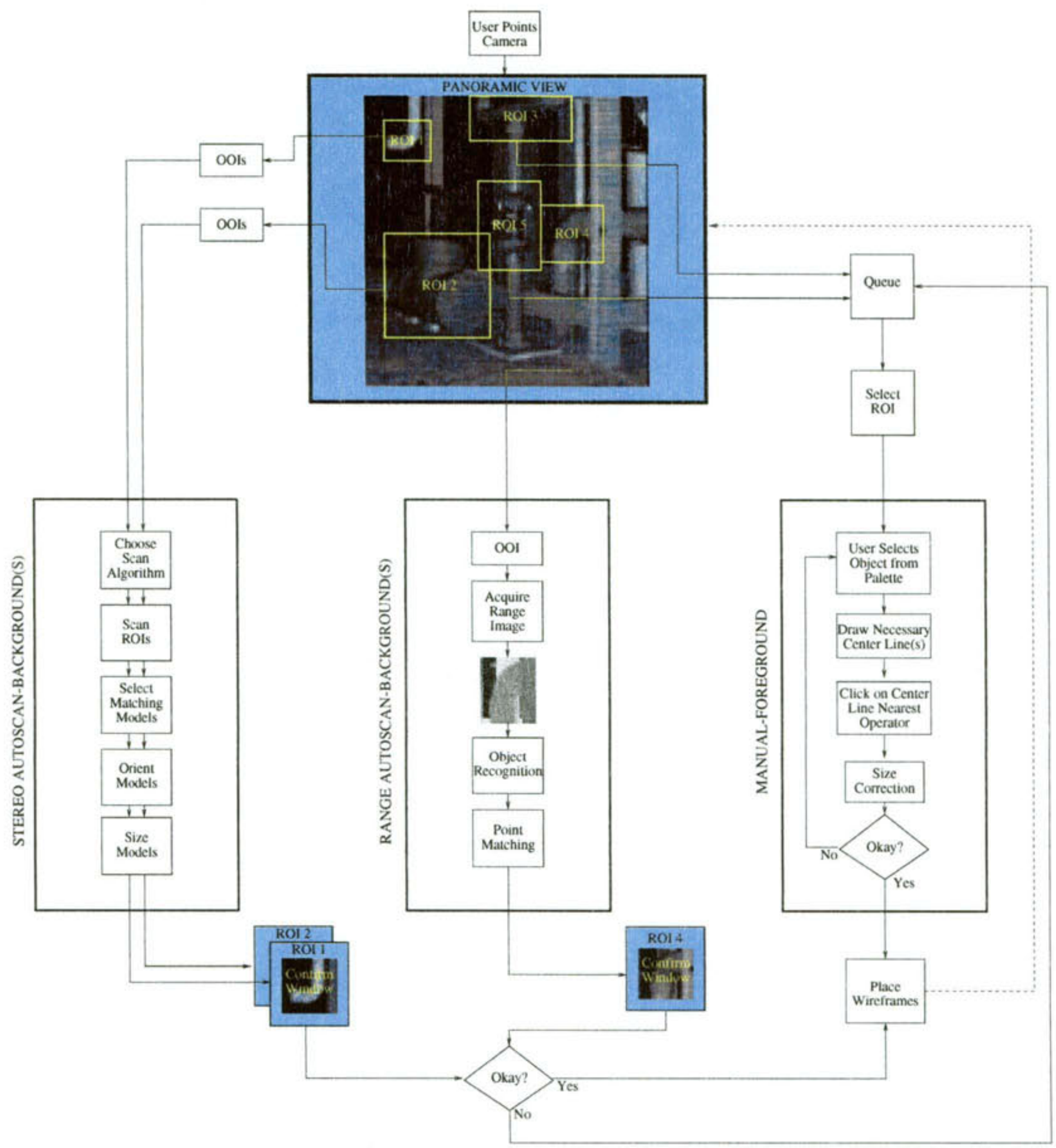

Figure 7-1: RTSA Functional Architecture

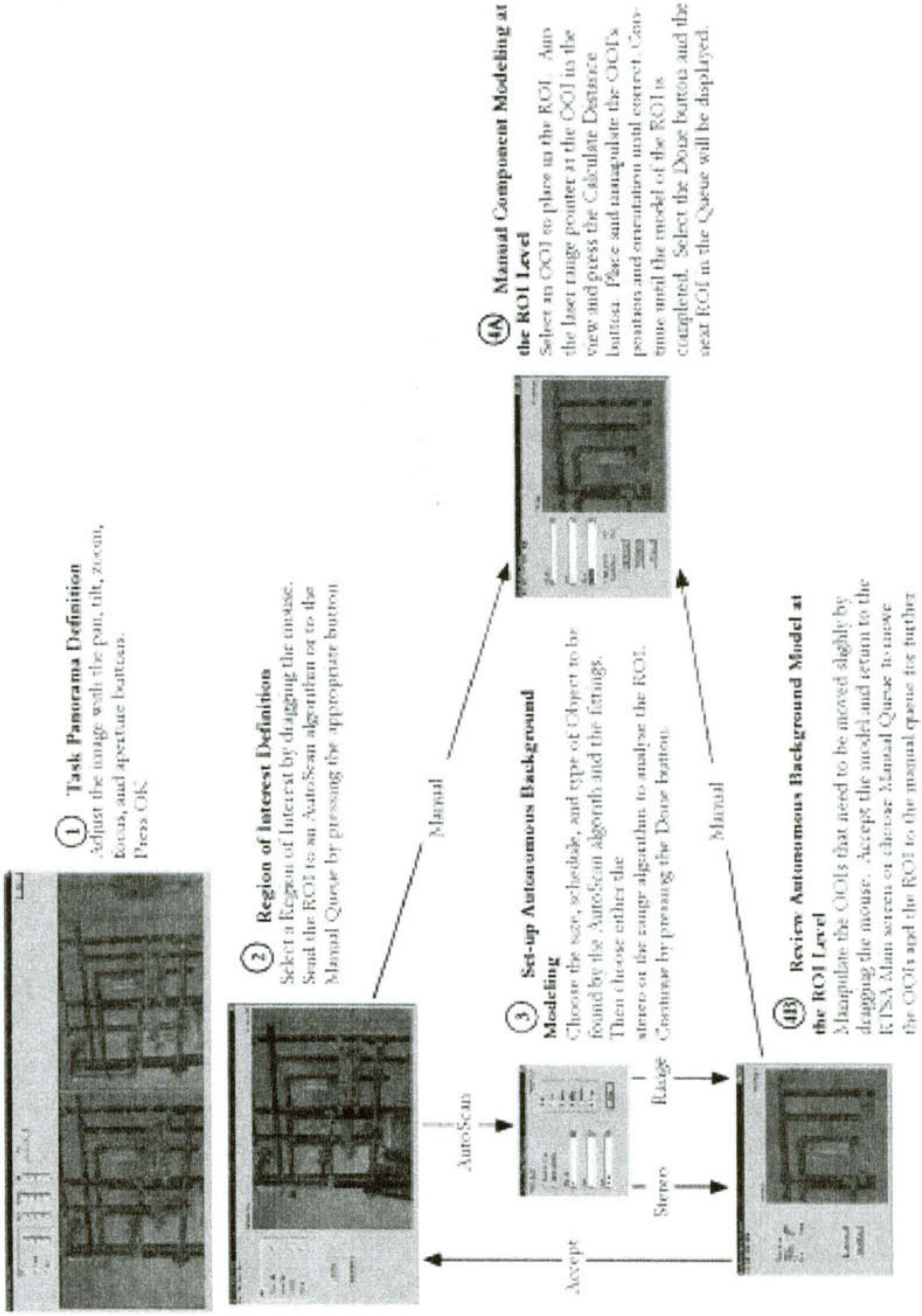

Figure 7-2: RTSA Operational Flow Diagram

7.1.1 Autonomous Object Recognition

To initiate the autonomous background algorithms, each ROI selected for AutoScan will be associated with objects of interest (OOI) to be modeled; information specified for each OOI will include details concerning parameters such as size, type, schedule, and material. Once each ROI has been selected and the information concerning component configurations has been supplied, the operator sends the ROI to one of the autonomous modules. The two autonomous modules available with the RTSA utilize either stereo or range information. The fundamental flow of operation for each of the modules is similar in that each module takes information from both the operator and the respective sensor head as an initiation for the object recognition process. The two modes of auto-recognition are discussed below. Section 7.1.1 discusses the stereo AutoScan procedure, and section 7.1.1 discusses the range AutoScan procedure.

Stereo AutoScan (Background Operation)

For each ROI selected for stereo AutoScan, the stereo sensor head is zoomed and focused automatically to accommodate the ROI as closely as possible at maximum zoom such that a close-up image can be obtained by the frame grabber. Using this image and the supplied component information, the AutoScan logically selects scanning algorithms for the object recognition process. Subsequently, the background scanning operations proceed and matching models are placed, oriented, and sized autonomously. When the AutoScan procedures have placed the recognized objects, the modeling result is brought to the foreground in a confirm window for examination by the operator. If the AutoScan process has trouble modeling the scene, then the AutoScan procedure would be executed in a timely manner; the procedure would be stopped at the timeout limit of approximately five minutes and the results displayed for operator review. When reviewing the model produced by the AutoScan process, the operator can accept, discard, or refine the any or all of the results. All accepted models are updated in the PV. If any work remains after the AutoScan process, the ROI must be sent to the manual queue for foreground manual analysis by the operator. In this manner, the effectiveness of the AutoScan modules is used to whatever extent available; any autonomous object placements - no matter how few - will reduce the overall workload of the operator and subsequently reduce the time required to produce the overall model.

Range AutoScan (Background Operations)

The Range AutoScan procedure is similar to the Stereo AutoScan in operational flow. The ROI and component information is sent to the module, and the SAR is used to autonomously obtain a range image. The background process is initiated and the module uses the supplied component and range information to perform point-matching analysis for object recognition. When completed, the AutoScan result is placed in a foreground pop-up confirm window for examination by the operator. Accepted models are updated in the PV; unacceptable results are sent to the manual queue for manual object placement.

7.1.2 Manual Object Placement (Foreground Operation)

Manual object placement is the most reliable and robust method for modeling OOIs since it is based on the operator's analysis of the scene. For the manual placement of object models, the operator selects a ROI from the manual queue. The selected ROI is subsequently represented in the foreground working window by the close-up frame capture that was autonomously obtained by the stereo head. Next, the operator selects an object from a hierarchical palette containing standard process components or selects custom to model an object in the task scene as a combination of primitive shapes using cylinders, rectangular solids, and spheres. Once the OOI is selected, the operator identifies points on the object in the task scene where the model is to be placed. By selecting two points for a pipe or three points for any other fitting on the object in the task scene image, the laser range pointer is used to obtain the distance to the object and calculate the desired position of the model. The operator's task is then to adjust the orientation and position of the translucent model until it coincides with the task scene view. By changing views to either of the stereo images, the operator can verify the position of the model and proceed to the next OOI. Once all the objects in the ROI are placed to the satisfaction of the operator, the model is accepted and the PV is updated as in the AutoScan cases. These procedures may be repeated as needed until each ROI in the PV is modeled as desired. For large OOIs, previously placed OOIs be used to define the end points of the large OOI; for instance, if a pipe extended from one ROI to another and the pipe was connected to an elbow in each ROI, then the elbow models could be selected as reference objects and a pipe model constructed between them.

7.2 Human Interactive Stereo System

The RTSA Human Interactive Stereo system is actually a multi-mode modeling system that is based on using the human operator interactively to optimize the overall model building speed of operation. Earlier R&D on task space scene analysis using stereo range imaging has been substantially expanded to include greater functionality, enhanced range accuracy, and background automation features that provide parallel autonomous region of interest analysis. Also, a robust and simple operator-based scheme has been incorporated to allow the operator to create task model objects directly on his screen when appropriate.

7.2.1 HIS Graphical User Interface

The HIS Graphical User Interface (GUI) is being developed with emphasis on communication efficiency. The HIS GUI's purpose is to allow the operator to intuitively interface with the computer to allow a natural and efficient flow of information to the computer.

Microsoft® Visual C++® 6.0 is being used to develop the GUI. The GUI is a Windows® 32 bit application based on the dialog box application. The dialog box was chosen over the single document or multiple document type application as it allowed the placement of images and data on the desktop in an orderly fashion. Dialog boxes also allow a more intuitive approach for displaying only that information that the operator requires while allowing requested information to be shown in another window.

The Matrox® Imaging Library (MIL) is used to access the frame grabber and is used as a statically linked library. The RTSA project is using a Matrox® Meteor PPB RGB frame grabber. The RGB frame grabber was selected since synchronized two black and white images are required at the same time. Currently, the red image is the left camera of the stereo pair and the green image is from the right camera; the blue channel is not used. The frame grabber requires several initialization steps before grabbing frames. The first step is to allocate resources for the application, system, and display(s). The next step is to inquire about a digitizer and allocate resource for it. Then buffers for storing the images need to be allocated. In the case of the RTSA project, three image buffers are required, one each for the RGB image, and the left and right black and white camera images. These steps define all the initialization that is required for operating the frame grabber; this code is implemented in the application's initialization function which is run when the application is started. The code responsible for grabbing frames is located in the dialog's initialization function and involves a call to the digitizer to grab frames continuously, another call to the digitizer to run a function to update the black and white image buffers from the RGB buffer at the end of each grab, and two calls to set the displays of the black and white images in the grab dialog's animate boxes. Once the required images have been obtained, the continuous grab is halted and the image buffers stored as image files in tiff format. When the application is done and no more frames will be grabbed, the application, system, displays, digitizer, and image buffers that were allocated are freed within Windows NT®.

The display of the model for displaying the PV and the ROIs is accomplished using Envision®. Envision® is used in the RTSA project to visualize stored models of standard process piping and to combine simple shapes in the construction of custom objects. The models of the pipes and fittings are stored in an Envision® file format and the scenes images are texture mapped onto large wall objects so that the image of the scene will appear behind the scene model. Unlike the MIL software, Envision® cannot be accessed with a statically linked library, Envision® must have an application running with the display in a RTSA dialog box window. When Envision® is displayed, an object is displayed behind the model with a texture map of an image; the image is either the left or right stereo camera view, the range image, or the intensity image. The texture map is required to give the operator information about where objects in the model need to be in relation to the real world. Since RTSA is a model-based approach, the task of Envision® is to display models of known objects; this allows the models of pipes and fittings to be constructed in Envision® before the scene is analyzed. When Envision® is requested to display an object, the corresponding file is recalled and the model placed at a specified position and orientation.

7.2.2 Foreground HIS

The foreground process is the HIS GUI which transfers information between the computer and the operator. Since the RTSA requires the operator to be an interactive part of building the model, the computer must display relevant information in a straightforward and uncluttered way. Therefore, the HIS GUI closely follows the architecture of the overall RTSA project as can be seen in Figure 7-2. The information that is first needed by the computer is the stereo images and the range and intensity images of the task area of interest; so the first screen that operator sees when starting the RTSA program is a live video view of what is coming from both cameras on the stereo head; this screen is called the Panoramic View Screen and is seen in Figure 7-3. The operator adjusts the pan-tilt unit and the servo lenses until the desired PV is in both stereo images and the images are in focus. When the operator is satisfied that the images show the PV correctly, then the OK button is pressed, and the next screen to appear is the RTSA Main Screen. In the RTSA Main Screen seen in Figure 7-4, the operator can view either the left stereo image, the right stereo image, the range

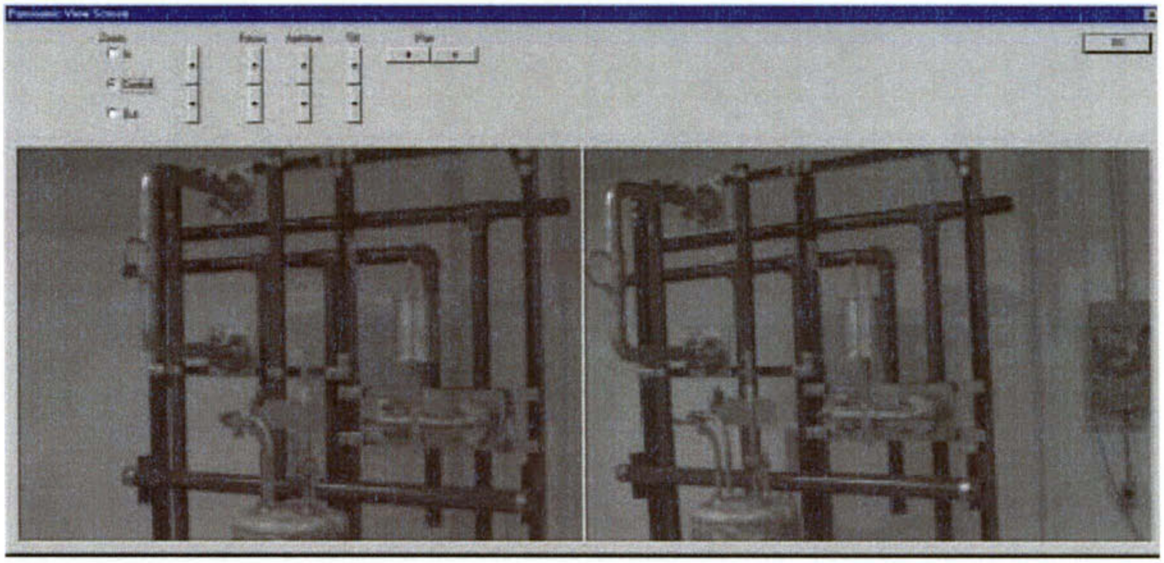

Figure 7-3: Panoramic View Screen

camera intensity image, or the range camera range image. The model is actually shown in this screen as well, but since no objects have been added to the model at this time, none appear. The RTSA Main Screen has access to all the functionality necessary to completely model the standard process components in the PV. The operator can select a Region of Interest (ROI) by dragging the mouse over an area in the PV; that ROI can then be sent to the Manual Queue where the operator manually models objects in the ROI, or the operator can direct the ROI to one of the background automated operations if he feels it is a strong candidate for autonomous processing. When the operator chooses to allow autonomous modeling within the ROI, the AutoScan Object Class Screen would appear as seen in Figure 7-5; in this screen the operator can choose which TSSA method to use, either the stereo TSSA method or SAR. The next step is for the operator to define the class of objects to be found. The class of objects is defined by schedule, nominal pipe size, and type; the type defines the style of connection used between parts as being either screwed, flanged, or welded. The last step is to select the objects to be found which are one or more of tees, elbows, pipes, flanges, and custom. Of course, flanges could only be selected if the type of pipe to be found were flanged.

Once the TSSA method algorithm has completed its analysis, the AutoScan Approval Screen will appear as seen in Figure 7-6; the purpose of this screen is to allow the operator to inspect and tune an Object of Interest (OOI) in the model created by an TSSA method algorithm.

The operator can approve or reject the model of a ROI. If approved, the OOIs are placed in the final model of the PV and can be seen in the RTSA Main Screen. If the model is rejected by the operator, the ROI is placed in the Manual Queue. As can be seen in Figure 7-7, the Manual Work Screen displays the current ROI to be modeled by the operator.

Through a series or pull down windows and menu selections, the operator can select an OOI to be placed in the model. The operator next has the distance the OOI needs to be placed in the model calculated using the stereo images and/or found with range information from the Sick Optic Electronic laser range pointer or SAR range images. Once the distance is known, the OOI can be

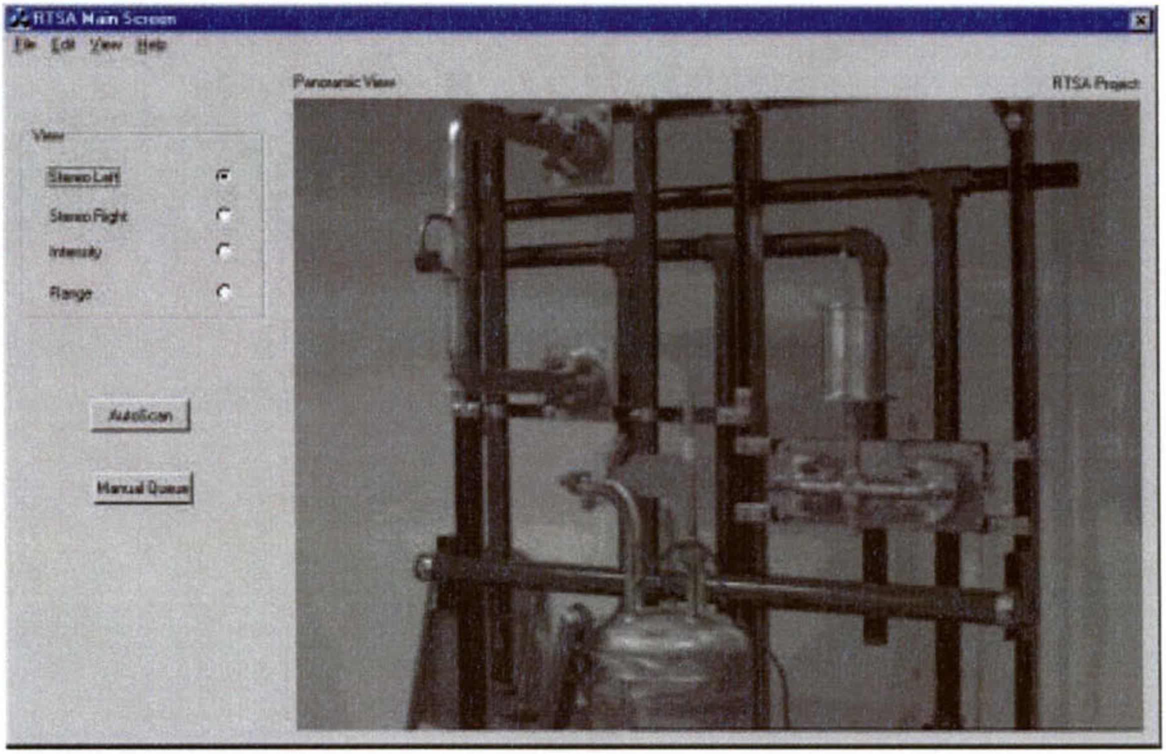

Figure 7-4: RTSA Main Screen

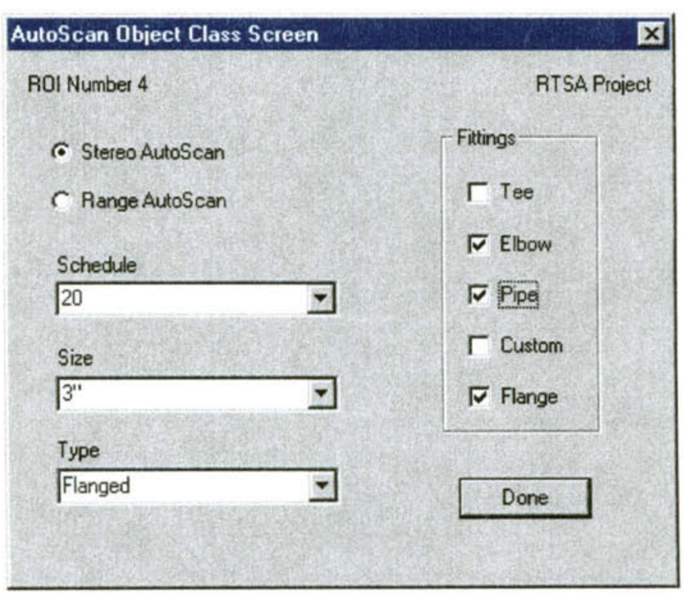

Figure 7-5: AutoScan Object Class Screen

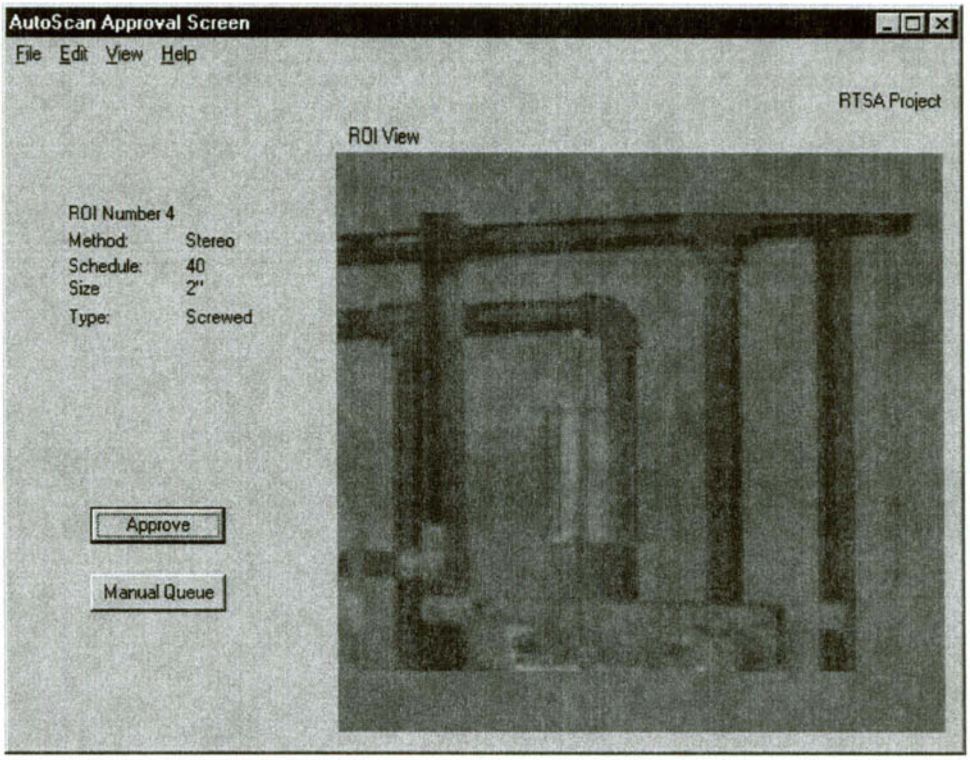

Figure 7-6: AutoScan Approval Screen

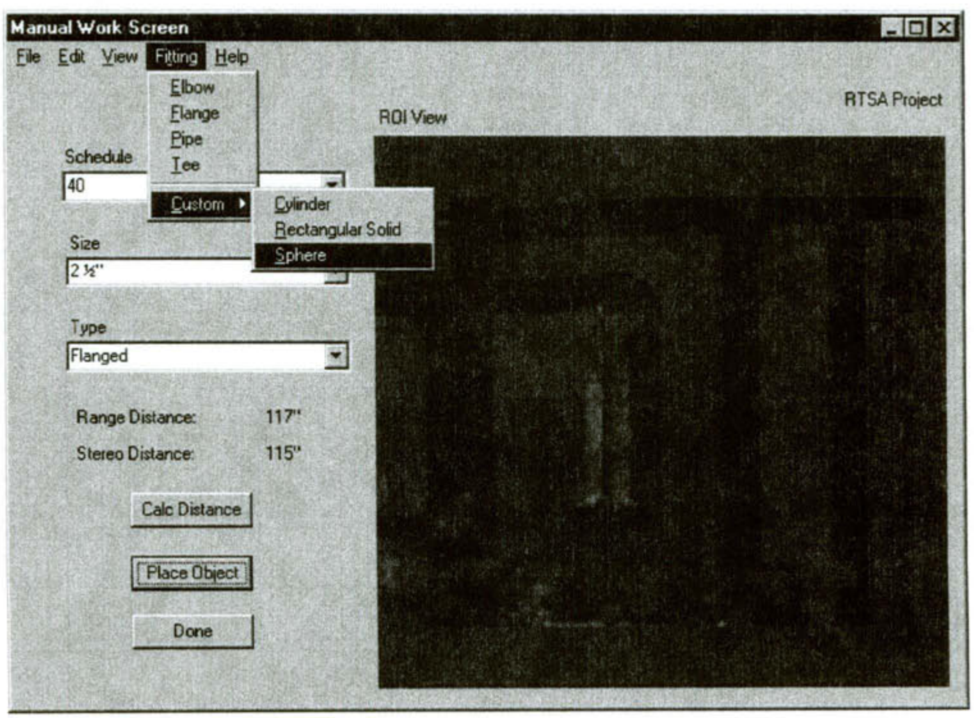

Figure 7-7: Manual Work Screen

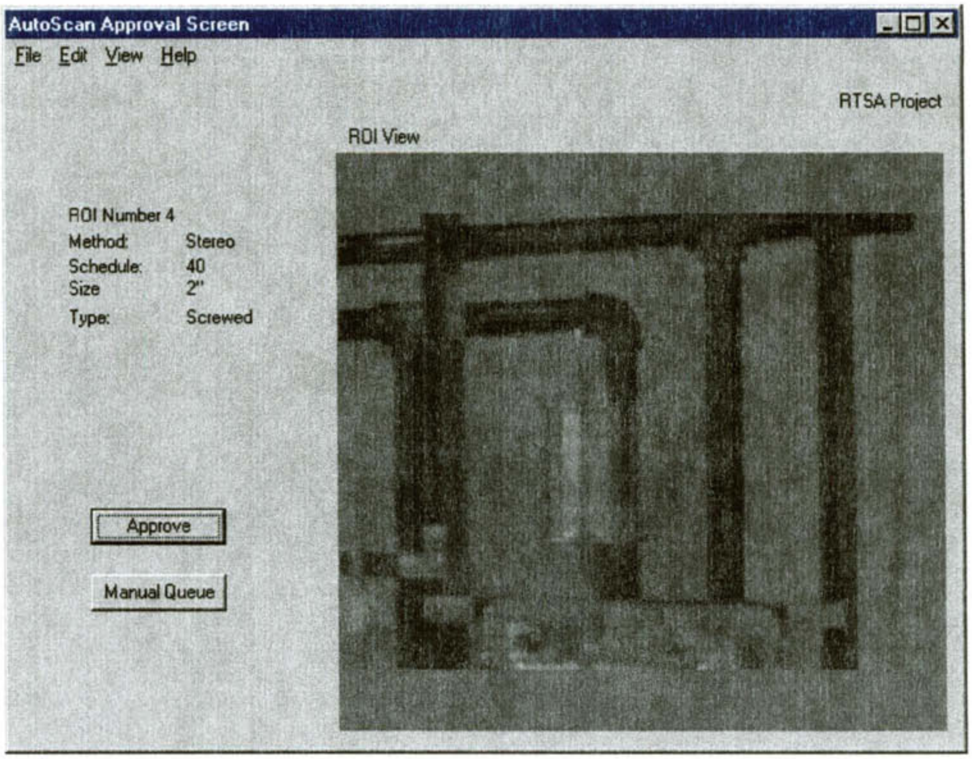

Figure 7-6: AutoScan Approval Screen

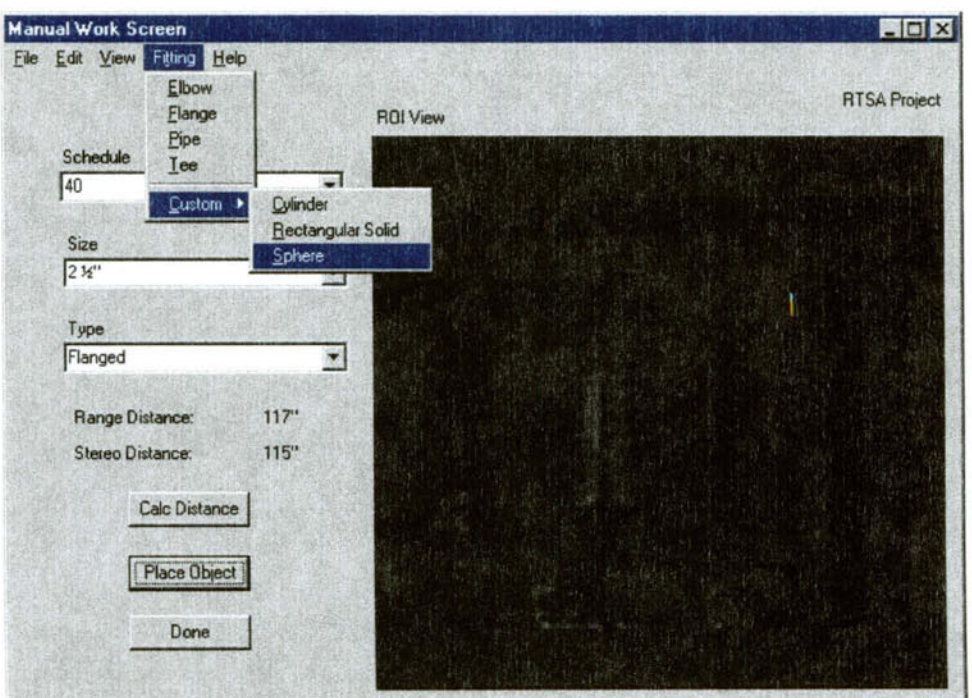

Figure 7-7: Manual Work Screen

placed in the model and then translated and rotated into its final position with a combination of mouse drags and keyboard inputs. OOIs are placed in the ROI until the operator is finished at which time the next ROI in the Manual Queue is displayed. Other information available to the operator is found in the ROI Status Screen where each ROI is listed with the method of evaluation and the status of that evaluation either approved, sent to manual queue, in progress, or queued to be done. Also, the information on the stereo head can be accessed in the Stereo Head Status Screen.

7.2.3 Background HIS

Currently, the system is equipped with auto-stereo algorithms specifically designed to determine the pose of pipes, elbows, and tees. Each of these algorithms, however, share the same five step approach: (1) edge detection, (2) object segmentation by edge shape extraction, (3) model-based feature correspondence, (4) stereo-based pose estimation, and (5) operator review.

Automatic Depth Measurement Using Model-Based Stereo

While the HIS system can be used to define accurate work space model poses for standard process components in the field of view, it is an operator-intensive process. Automatic stereo methods can help reduce the operator workload, but the development of automated techniques that are robust to both photometric and geometric variations have proven quite difficult.

However, three key properties of the RTSA work environment made the incorporation of automatic stereo processes attractive. First, the majority of standard process components in the RTSA system's work space are comparatively simple structures such as pipes, elbows, tees, flanges, and bolts. Since these objects have geometries that can be specified using a small number of parameters, it is more tractable to determine their locations in visual imagery using automatic segmentation techniques, assuming the view-points are not degenerate. Second, since exact dimensional specifications are available for these objects in the scene, model-based techniques can be used to greatly simplify the stereo matching process. Third, since these object types are the most common, automatic algorithms for locating them in images need not be perfect. For example, if only 70% of the pipes in a given image are correctly recovered, this would still significantly reduce operator work load.

A number of object-specific, automatic stereo algorithms, or AutoScan algorithms, in the RTSA system that can be used interactively. To use these procedures, the operator first chooses a ROI, and specifies a class of objects. The auto-scan procedure then executes as a background process while the operator is free to perform other tasks, such as using the HIS to specify the location of more complex vessels in the scene. Once completed, the results of the auto-scan procedure are displayed in a window. In this window, the recovered object models are shown as partially transparent overlays on the original stereo pair. The operator is then given a chance to edit the results by removing spurious model placements and by adjusting the location of slightly misaligned models. Once the operator is satisfied with the results, the locations of the objects are transferred to the workspace environment model where the model shown in the PV is updated.

Edge Extraction and Object Segmentation

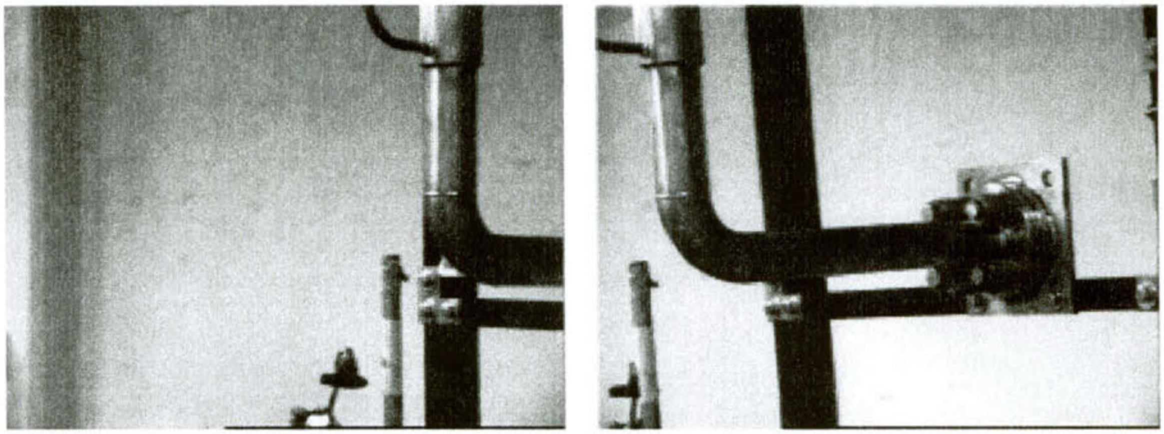

Figure 7-8: Stereo Image Pair

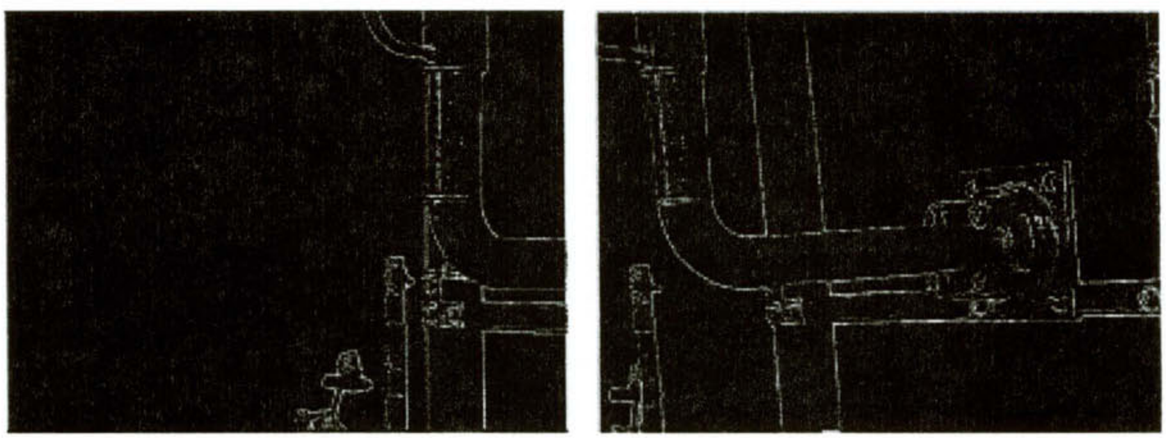

Figure 7-9: Laplacian of Guassian Stereo Image Pair

As demonstrated in Figure 7-8, there is typically significant intensity contrast between the process components and the task space background which suggests using edge-based segmentation algorithms in our auto-stereo routines. For edge-detection during pre-processing, the Laplacian of Guassian (LoG) [1] operator is used. Appropriate values for the standard deviation and mask size were determined empirically through experimentation with numerous stereo image pairs of the laboratory testing mock-up under varying lighting conditions. An example of a typical LoG image pair is shown in Figure 7-9. Since the Hough transform only returns the functional form of the prominent image curves, connected component analysis is performed in the neighborhoods of these curves to determine their exact image length and location. Any lines curves with lengths less than a user-defined threshold are then discarded as noise.

For pipe extraction, straight lines are extracted using the Hough transform. Pipes are then identified through line pairing. Any two lines that are approximately parallel and are approximately of equal length are labeled as potential pipes and placed as a pipe model in the ROI.

For elbow extraction, the Hough transform is use to extract parabola-like objects. Elbows are then identified through arc pairing. Any two arcs that have an apparent common center consistent

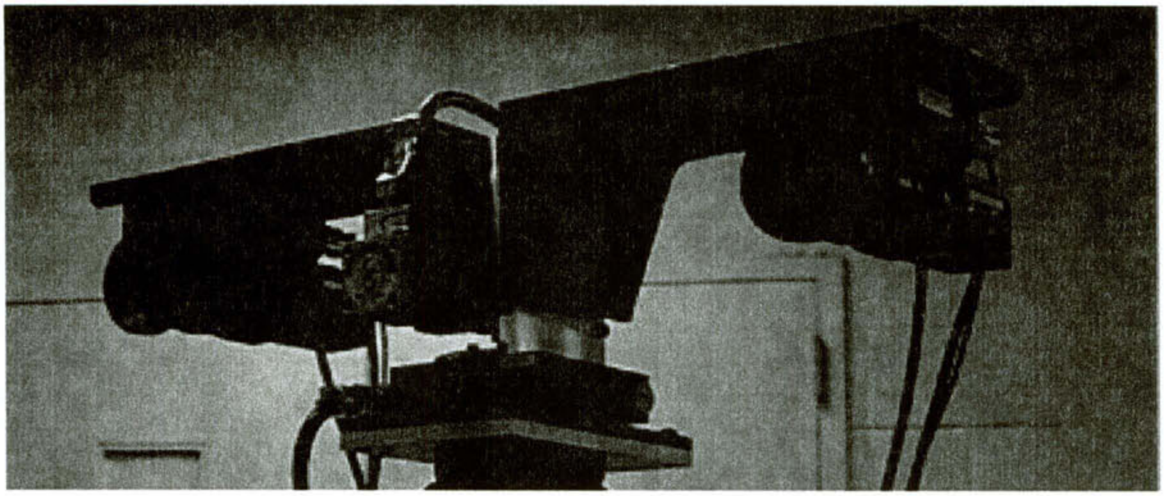

Figure 7-10: Stereo Vision Sensor Head

with the standard components elbow library are labeled as potential elbows.

Straight lines are again extracted for tee segmentation. Potential tees are identified through an investigation of line intersections. Since pipes are circularly symmetric, the intersection of two orthogonal pipes will always appear rectangular in the edge image. Thus, potential tees can be found by examining the intersection of lines recovered using the Hough transform. Any four lines that intersect to form a roughly rectangular region with an area less than or equal to a pre-defined threshold are labeled as potential tees. This threshold is related to the size of the pipe.

7.2.4 Stereo Vision Sensor Head

The sensor head for the RTSA HIS consists of a pair of black-and-white CCD video cameras mounted to a combination pan-tilt stage as shown in Figure 7-10. Each camera is fitted with a servo-actuated zoom lens. Both the pan-tilt and zoom lenses are controlled via a standard serial interface. The composition of control software for the pan-tilt unit and the servo-lenses is complex and has involved significant development effort. The first step in the development of this software was the assembly of information pertaining to the operation of serial ports and serial communications. It was found that the methods required for both control on the serial ports and the transfer of information via serial communications is highly computer platform dependent. Due to this, it was decided to build the control software in a modular fashion. The bottom layer would provide an interface between a user level program and the details of serial port operation, e.g., opening of a serial port, setting of communication parameters, reading/writing data to/from the port. The top layer provides a set of simple function calls which allows the control of the pan-tilt unit and the servo-lenses for integrated HIS operation. The major advantage of this approach is that it makes porting the software to other computer platforms relatively easy since only the bottom most layer, i.e. basic serial port control/communication, would have to be rebuilt. At present, a base class for serial communications under Microsoft® Windows NT® has been written.

The top layer of control functions for control of the pan-tilt unit consist of motion functions and status functions. The motion functions allow the RTSA operator to maneuver the pan-tilt head in both a relative, i.e., move the pan axis 10° CW, and absolute, i.e., move the pan axis to 30° CCW.

For the specification of an absolute movement to be meaningful, the definition of a "home" position is required. Towards this, a function was constructed which takes the current pan-tilt position and assigns it as the origin of its coordinate system. Also included among the motion functions are ones which allow the operator to set soft travel limits. These provide the ability to limit the region of interest and can be utilized to prevent lost motion due to camera movement beyond the region of interest. Status functions give the user some feedback as to the current state of the pan-tilt unit. Information such as current position, axis motion, and proximity to soft limits are available. The determination of the current position of the pan-tilt unit is of utmost importance to the stereo calculations. Since the RTSA has both interactive and autonomous modes of operation with only a single pan-tilt unit, axis motion information is needed for scheduling purposes. This information can be used to prevent any two processes from trying to control the pan-tilt unit simultaneously by blocking all other requests until the current motion is complete. Lastly, the detection of soft travel limits allows one to communicate to the user that one of the travel limits had been reached and therefore all subsequent moves in that direction will be blocked. Currently, all the above functions have been written and work is underway to complete a graphical user interface for the pan-tilt unit.

The servo-lenses allow computer control of three basic parameters: focus, zoom, and aperture. Functions for controlling and monitoring these parameters are similar to those presented for the pan-tilt unit. The main difference is in the command syntax. Independent control of any of the three basic parameters is implemented in a set of simple function calls which commands each lens to change to the requested value. This precludes the need for software monitoring of such events. However, monitoring of other parameters is required for the purposes of user feedback and system functionality. Functionality for continuous monitoring of zoom, focus, and aperture values as well as the detection of ring motion was deemed necessary. Knowledge of the current zoom setting, i.e., focal length, is a necessary parameter for stereo calculations. The detection of ring motion, i.e., movement of zoom, focus, or aperture, is utilized for scheduling purposes necessitated by the parallel nature of the RTSA, i.e., autonomous and interactive processing, by preventing two tasks from accessing the lens at the same time.

The stereo vision head will also include a single point laser ranging device for the purpose of directly measuring the range to a designated point on an object of interest. The laser pointer is used as a redundant measurement in stereo camera calibration and in the manual model construction mode of HIS.

7.2.5 Stereo Platform Calibration

The RTSA stereo camera platform is calibrated using a variation of Tsai's original calibration algorithm [2]. However, since the stereo vision head is equipped with a visible laser range pointer, no calibration target is required. A great advantage of this layout concerns in-situ calibration; i.e., this system can be recalibrated as needed while remaining deployed in the task work area.

When the calibration procedure is executed, the stereo platform is moved back to its home, or nominal, position. The laser is then activated and the apertures of the cameras are adjusted so that the laser range pointer "dot" is easily detected in the scene using simple gray-level thresholding techniques. The range value from the laser is then recorded, along with the image positions of the laser dot. The platform is then moved to a random pan/tilt position. The range return from the laser is recorded again, along with the dot locations in the image pair. This procedure is repeated until a minimum number of unique range values have been recorded. Tsai's calibration algorithm

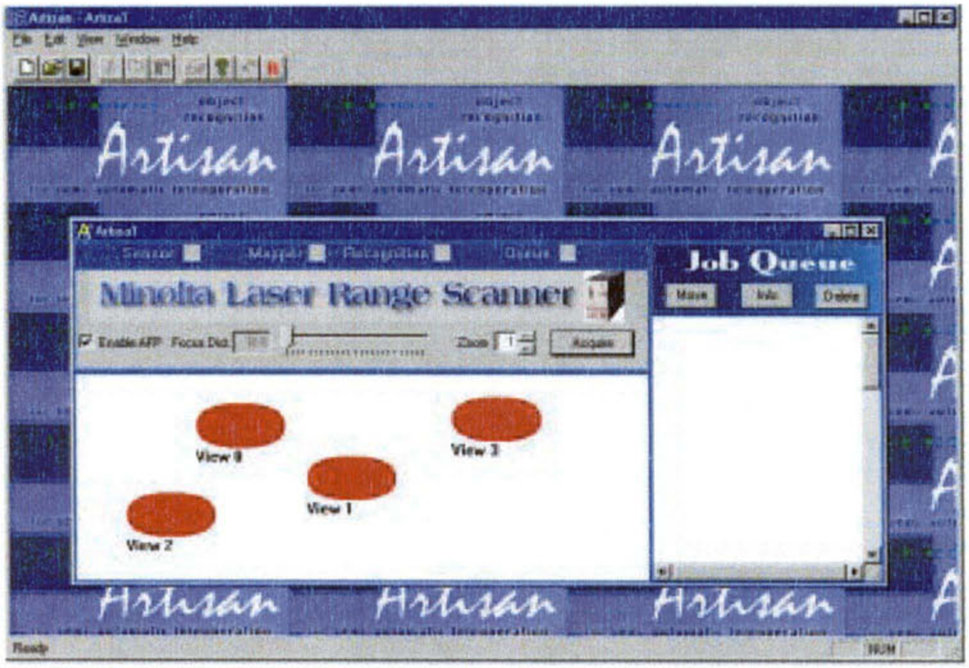

Figure 7-11: New Interface Window for Range Image Merging

is then employed to recover the rotation and translation of each camera relative to the laser range pointing system, along with their focal lengths.

To help reduce the effects of noisy measurements, the system continues to move to new positions, obtain additional range points, and re-compute the solution until either stopped by the operator or until the rotation and translation parameters reach steady state values for both cameras taking approximately twenty to thirty points. At each re-computation stage, the new estimates for relative rotations and translations between cameras and range pointer are displayed so that the operator can evaluate the process.

At the conclusion of this process, the calculation of the relative rotation and translation between the two cameras is straightforward given the transformation of each camera with respect to the laser range pointer.

7.3 Semi-Autonomous Range Analysis System (Artisan)

7.3.1 Graphical User Interface Development

A new version of the original Artisan user interface was developed for operation on personal computers (the earlier version having been restricted to operation on Silicon Graphics® workstations). All functionalities have been preserved and a new one for merging range images has been added. The operator will also be able to select several objects from the model catalog for simultaneous object recognition, display multiple images, view multiple meshes and combined meshes. Integration with the PC version of Envision® has been deferred to the optional integration phase of the project. Two screen captures are shown in Figures 7-11 and 7-12.

Human factors engineering and analysis has been used to both craft and optimize a unified

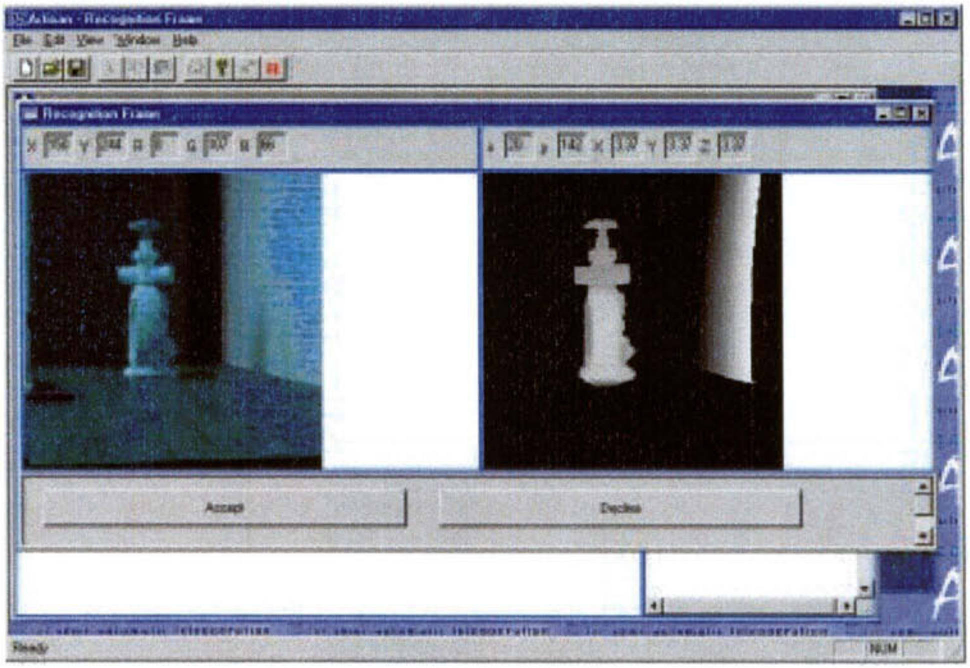

Figure 7-12: New Interface Window for Display of Range Data

GUI - SAR user interface. Several prospective screen layouts that accommodate the flow of actions have been designed and evaluated. Window and menu structures, iconic representations, use of standardized object models, and graphical overlays have been studied in detail. Software to implement a working prototype has been developed. The GUI - SAR has been ergonomically optimized through an iterative process.

7.3.2 Range Image Analysis

Building upon the spin-image matching technique developed previously for free-form object recognition, two new scene analysis tools have been developed. In the course of developing these new tools, several additional insights to the behavior of the algorithms themselves have been applied to improve object recognition performance. These new tools are direct extensions of the range image analysis technique developed for recognizing single objects. That fundamental technique uses the concept of spin images as a way to represent and compare collections of 3D data. A brief review of that technique follows.

The spin-image is a two-dimensional descriptor of the local shape of a free-form three-dimensional surface at a point p on that surface. Spin-images encode the positions of points near p in terms of distance along and distance from the approximated normal to the surface at p. A measure of local shape similarity between the surfaces surrounding two points is obtained by comparing this coordinate information from the spin-images at these two different points. To construct the spin-image for an arbitrary point p, the best-fit plane to the nearest neighbors of p and approximate the normal to p as the normal to this plane. After this a 2D basis using the normal n and the plane P perpendicular to n and passing through p is defined. This process is illustrated in Figure 7-13. The (a, b) values are then discretized and accumulated into a 2D array of bins called a spin-image;

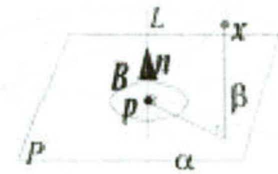

Figure 7-13: Oriented point basis for spin image formation.

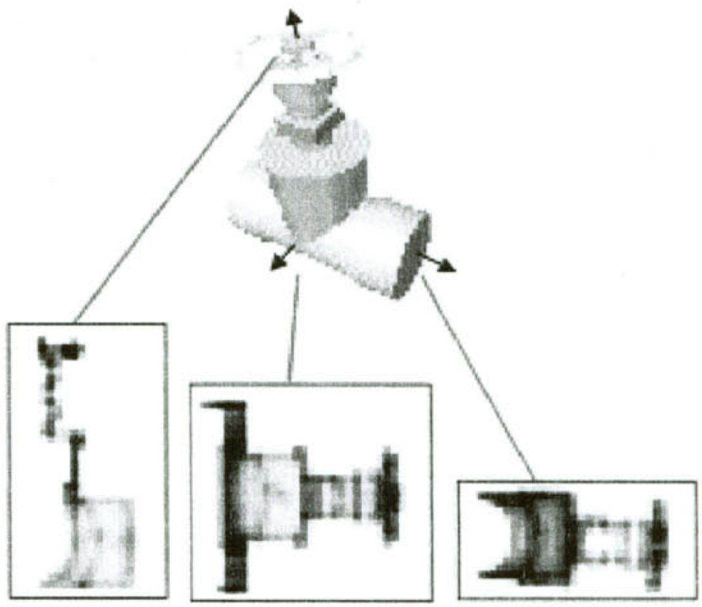

Figure 7-14: Spin images of a typical valve.

each bin in the spin-image corresponds to some range of a and b values. These spin-images are compact descriptors of the local shape of a surface around a particular point; if two points have similar spin-images, they are considered to have similar local shape. Three examples of spin images are shown in Figure 7-14.

By repeating this process for all points in a collection of 3D points, a stack of spin images is generated that serves as an alternative description of the 3D surface.

Spin-image matching provided a powerful tool for matching data acquired from range sensor scans at unspecified locations to models of objects that are expected to be found in the scene. Regardless of the sensor used, range data from it consists of 3D points in some fixed coordinate frame. Given these points, a triangular mesh is formed by connecting nearest neighbors, and noisy points and edges in this mesh are removed through cleaning and smoothing operations. Now a high-resolution 3D representation of the model as seen from a particular viewpoint is available; however, this can be simplify this to obtain a low-resolution version of the surface.

A plausible transformation from model to scene is calculated from each group of correspondences using the algorithm of Horn. Verification of transformations is performed by transforming the model surface mesh into the scene. Then for each model vertex, its closest scene vertex in the six dimensional space of vertex positions and surface normals is determined. If a large number, (e.g., one third the vertices in the model) of model points and corresponding closest scene vertices have a closest distance that is less than two times the mesh resolution of the model, then the transformation is verified because it brings a large number of scene and model points into alignment with one another.

It should be noted that the range algorithms could be improved through the following enhancements:

- Reducing Effects of Clutter and Occlusion
- Simultaneous Recognition of Multiple Objects
- Merging Multiple Range Data Sets

However, even without these enhancements the technique was successfully applied in the RTSA system and was able to map a mockup of a large scale facility.

7.4 RTSA Hardware Architecture

7.4.1 Overall System

The overall RTSA system consists of a dual 333 MHz Pentium II® PC running the GUI, a dual 300 MHz Pentium II® PC connected to the stereo head, and a Silicon Graphics® connected to the laser range camera. The 300 MHz PC controls the stereo head via four serial ports connected to the pan-tilt drive, each of the two servo lenses, and the laser range camera. Input from the two, black and white, CCD cameras is transmitted via the red and green channels of a standard BNC cable system. The 300 MHz PC captures the images using a Peripheral Component Interconnect, RGB frame grabber. The LASAR laser range camera is connected to a VME rack; this VME, whose program is compiled on a Sun® workstation, controls the LASAR. The SGI communicates with the VME and also has a frame grabber for capturing images from a camera. Both the 300 MHz PC and the SGI are connected via Ethernet to the 333 MHz PC running the GUI. The computations involved in the AutoScan algorithms are completed on the SGI and the 300 MHz PC so that response to the operator's input is not delayed. Conversely, the AutoScan tasks can be completed more quickly since the SGI and the 300 MHz PC do not have to maintain and update a display which is computationally intensive. A schematic diagram of the computational hardware layout is shown in Figure 7-15.

7.4.2 Physical Arrangement

The physical arrangement of the RTSA laboratory testing system is shown in Figure 7-16. The stereo vision head and the laser range scanner are mounted on separate tripod stands. The elevation and standoff distances of these tripods with respect to the task mockup approximate the distances that would be expected in an actual mobile robot work system and are ten to twelve feet from the task scene mockup in this model. The locations of the head and scanner can be varied to provide different object perspectives and the task mockup itself can be rotated and relocated to change views and perspectives as well.

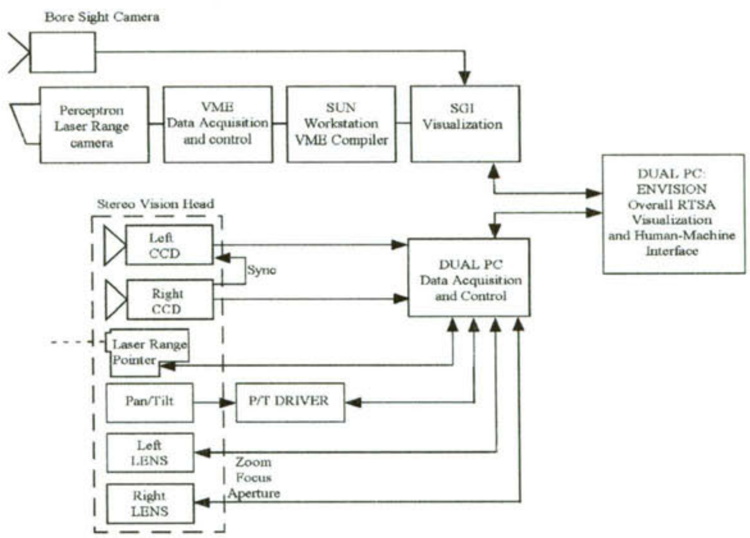

Figure 7-15: RTSA Computational Hardware Layout

7.5 Laboratory Testing and Evaluation

Based on earlier work and the general state of the art in range sensing, the conceptual merits of RTSA are unquestionable. The fundamental questions remaining are quantitative in nature and pertain to the achievable model building accuracy and speed. The laboratory testing task of the project is intended to address these quantitative factors in a rigorous manner. The following sections describe the overall laboratory test plans.

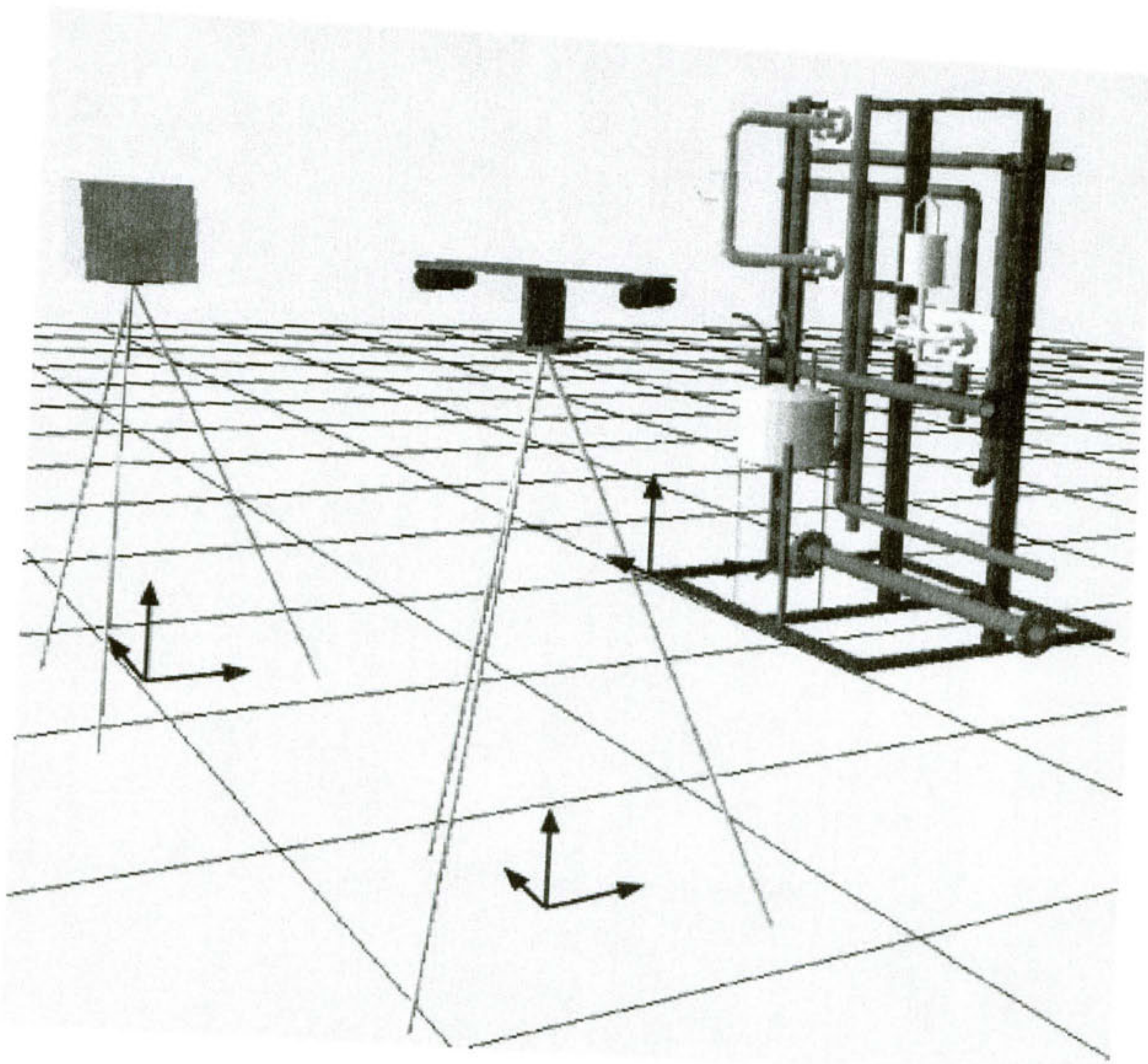

Figure 7-16: Laboratory Test Setup Arrangement

Experimental Method
The experimental objectives are two fold:

1. Evaluate the accuracy and precision of the RTSA geometrical measurements including
   - The percentages of time modules correctly recognize objects
   - The percentages of time modules fail to recognize or report wrong matches
   - The accuracy with which modules report dimensions, location and orientation of recognized objects
   - How fast the modules work
   - What pathological scene conditions result in failures
2. Determine key human factors issues and parameters.

The basic experimental approach is to perform repeated measurements of selected process equipment objects in the task mock-up and to compare the measured results with the true dimensions, position, and orientation of the particular component. The true position of selected components will be based on the task mock-up design data and independent survey measurements made of the task mock-up within the laboratory. The independent survey data will be referred to the coordinate reference system used for RTSA measurements for convenience. The experimental protocol will be defined such that all of the foreground and background scene analysis features of RTSA are exercised. Data pertaining to the amount of time required to construct each component model will be captured and multiple trials of all model building tasks will be performed to encompass learning effects. Multiple operators will be used to perform all of the tests. Operator errors and behaviors will be monitored and recorded during tests as well.

Task Mock-Up
The laboratory task mock up has been constructed from standard process piping and components to provide a realistic task space scene in terms of relative sizes, shadowing, occlusions, and surface characteristics. The mock up is shown in Figure 7-17. The unit has been constructed as a free standing floor module so that it can be repositioned easily for different sensor perspectives. As mentioned earlier, one of the key strategies of RTSA is to recognize that process piping and equipment for the most part is comprised of standard sizes and schedules. The laboratory task mock up incorporates a fairly large number of standard components as summarized in Table 7-1.

In addition, it incorporates several custom designed stainless steel remote components that were obtained from the Robotics and Process Systems Division at ORNL. The surface conditions of the process components have not been treated or prepared in any special way to ensure that the various sensors are dealing with realistic colors, contrasts, and energy reflections characteristics. The floor and rear wall aspects of the test area also provide realistic task visual characteristics. The floor is unpainted concrete and the masonry rear wall is painted and includes wall-mounted electrical boxes conduits.

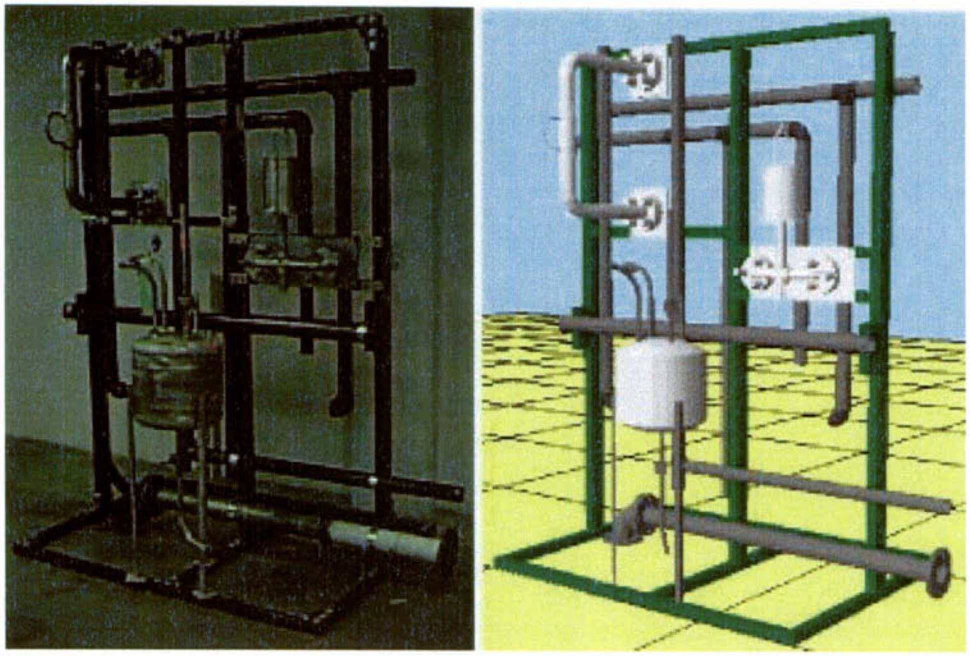

Figure 7-17: Laboratory Task Mock-up (a) photo (b) model

Table 7-1: Task Mockup Standard Process Components

Component Type	Schedule	Size (Nominal)	Connection	Number
90° Elbow	40	1 inch	Welded	2
90° Elbow	40	2 inch	Threaded	4
90° Elbow	40	2-inch	Welded	2
90° Elbow	40	3-inch	Flanged	1
90° Cross	40	1-inch	Welded	1
Pipe Tee	40	2-inch	Threaded	4
Pipe Flange	40	1-inch	Welded	2
Pipe Flange	40	2-inch	Welded	2
Pipe Flange	40	3-inch	Threaded	2
Straight Pipe	10	2-inch	N/A	14
Straight Pipe	10	2.5-inch	N/A	1
Straight Pipe	40	3-inch	N/A	1

Modeling Results Calibration

A theodolite digital survey system will be used to survey the positions of the task mock-up, HIS tripod and the SAR tripod within the laboratory. These positions and orientations will be permanently landmarked on the laboratory floor to ensure that equipment reference positions are invariant. During experimental campaigns, the reference positions will be periodically verified using the theodolite system. These survey data will define the geometrical relationship between the mockup and its components (via the task mockup design data) and sensor tripods which can in turn be used to calculate the true position and orientation of the individual process objects of interest.

The purpose of RTSA is construct accurate geometrical models relatively quickly. The laboratory testing will be designed to evaluate metrics that address this fundamental purpose. The most straightforward metrics relate to how accurately a specific component is modeled in terms of:

- process component type
- size
- position accuracy
- orientation accuracy
- component match up with adjacent/connected components

Human factors metrics will be more subjective in nature and will seek to evaluate operator performance, error rates, and fatigue factors.

8 Conclusions

A RTSA system which meets the objectives set forth in the purpose and shows that the hypotheses is valid. A key objective in the RTSA project was to use earlier work in task space scene analysis as a foundation for the development of an in situ geometrical modeling system which is a practical tool that typical remote equipment operators could use and feel comfortable with. Care has been taken in the design of the human-machine interface to assure its simplicity and ease of use.

The experimental evaluation of the RTSA system showed that the computer assisted telerobotic system was extremely beneficial in reducing the execution time associated with the control of a manipulator along a curved path. In addition, the results of these experiments show that the implantation of the algorithm tested resulted in a significant increase in task execution efficiency as shown by the reduction in mean execution time, frequency/magnitude of collisions, and reduction in operator fatigue. Thus, the experiment shows that an assisted telerobotic strategy maybe a viable method for conducting precise telerobotic operations in a variety of operating conditions ranging from radioactive cleanup to high-precision manufacturing to potentially life-saving surgical procedures.

References

[1] D. Marr and E. Hildreth. Theory of edge detection. *Proc. R. Soc. Lond.*, B207:187–217, 1980.

[2] R.Y. Tsai. A versitile camera calibration tecknique for high accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*, RA–3(4):323–344, 1987.

UNIVERSITY HONORS PROGRAM

SENIOR PROJECT - PROSPECTUS

Name: Surya P. N. Singh

College: Engineering Department: Mechanical, Aero., & Engineering Sci

Faculty Mentor: Dr. Hamel

PROJECT TITLE: A Smart Move

PROJECT DESCRIPTION (Attach not more than one additional page, if necessary):

Please See Attached

Projected completion date: June 1, 1999

Signed: ____________________

I have discussed this research proposal with this student and agree to serve in an advisory role, as faculty mentor, and to certify the acceptability of the completed project.

Signed: William R. Hamel, Faculty Mentor

Date: 5/9/00

Return this completed form to The University Honors Program, F101 Melrose Hall, 974-7875, not later than the beginning of your last year in residence.

Project Description

The extensive use of robotics in hazardous environments has been limited by the inability to easily model and control robotic systems with high accuracy. In general, purely manual robot modeling systems lack the precession needed and fully autonomous systems are unable to cope with high levels of environmental variability or abnormalities. It is the objective of this research effort to develop and use a computational modeling and control system consisting of a series of sensors and algorithms that will augment manual operations to not only minimize error, but to reduce the tedium and difficulty presently characteristic of this form of robotic operation.

The basic methodology of this system is to extend the traditional master-slave telerobotic model via a series of sensors coupled to a computational model of the environment, which together adjust the velocity scaling of the end-defector. In particular, this system is implemented using a set of transformations that adjust the velocity scaling of the end-defector in proportion to the distance from the end defector to the modeled object or location.

To design and test this system an experimental setup consisting of a multi-component mockup wall, a Titan robot arm, and a mini-master joint space controller. The experiment measured the time it took to select four rectangular arranged points on the surface of this wall. System performance was evaluated on the time necessary to accurately position and navigate the end-defector around the four points and the number of collisions and errors occurring throughout the process.

The results of these experiments show that the algorithm tested resulted in a significant increase in task execution efficiency as revealed by the reduction in mean execution time, frequency/magnitude of collisions, and reduction in operator fatigue. Thus, the experiment shows that an assisted robotic modeling and control strategy maybe a viable method for improving operator interactions with the robot system.