Chapter 1

Introduction

Localisation is a fundamental issue in mobile robotics, since most applications of autonomous robots moving in any type of environment, require a posture estimate to fulfil their mission. This dissertation introduces a novel solution for localisation of wheeled mobile robots based on partially known maps and partial environment descriptions computed from data samples acquired with different sensors.

Wheeled robots operate most often in planar surfaces and, under this restriction, the localisation problem is defined as computing the co-ordinates of some fixed point in the robot on the horizontal plane and its orientation respective to some known reference. To localise the robot an external spatial reference is required. This may be provided by numerous techniques, such as artificial beacons or landmarks, pre-existent environment maps, motion integrators. In addition, sensors are required to perceive the environment around the autonomous robot. There are many types of sensors, based on different working principles. However, and as a design option, the choice of particular sensor devices for localisation purposes is often constrained to the ones already present to fulfil other tasks.

The proposed localisation system was designed as a part of the RESOLV project, a European Union project aiming at creating realistic and accurate three-dimension (3D) models of structured indoors environments. The localisation algorithms are an element of the wheeled autonomous robot developed to perform this task. The RESOLV system applications range from the construction industry to the archaeology, virtual museum tours and many others. The sensors embarked on the robot are a laser range scanner, a monocular video camera, an ultrasound ring and an odometry measuring system. All these are available in the general configuration. In addition to this configuration, simpler variations were designed and these excluded the ultrasound ring and the odometers. Therefore, the localisation system should depend on the laser camera only, since the video camera is less adequate to distance measurements. The external map reference is computed from the 3D reconstructed model although it may be supplied by an architectural map like the Computer Aided Design (CAD) two dimension models or a digitised floor plan.

The localisation problem may be stated as: "How to compute the position $(\mathbf{x},\mathbf{y})$ and orientation $(\mathbf{\theta})$ estimates of a wheeled autonomous robot in a piece-wise horizontal ground using Laser Range data and odometry data when the latter is available. The world

reference is based on a partial 3D surface description, which is incremented as the robot progresses on the scene".

The main novelty of the proposed approach is that two requirements of previously known solutions were removed: no *a priori* posture estimate is necessary and the world map is usually unknown beforehand. To localise the robot it is necessary to autonomously identify and use natural features and resolve ambiguities and deadlocks in an automated manner.

The proposed localisation solution combines three modules. The first module is a feature extraction and matching algorithm with no a priori posture estimate. The second module defines a likelihood measure to compute the accuracy of the posture estimates. The third module uses co-ordinate transforms and a recursive loop to move from an initial posture estimate to a final enhanced estimate. This compound approach overcomes most pitfalls of individual algorithms, increases robustness, while keeping the algorithm development and analysis simple.

The dissertation is organised as follows: Chapter One introduces the subject and the working environment which integrates the proposed localisation algorithm. The focus moves from the general overview of the RESOLV project to the particular aspects of the localisation solution. Chapter Two covers the first algorithm based on feature extraction. This algorithm is the main novelty of the dissertation. Chapter Three discusses the likelihood measure, which is a statistics designed to assert the accuracy of the posture estimates computed by localisation. Chapter Four covers the second algorithm based on an initial posture estimate, which is refined by small steps while using the likelihood measure to assure the convergence to a more accurate solution. Chapter Five discusses the overall performance of the algorithms and presents global experimental results, under different operation conditions. Chapter Six presents the conclusions and directions for further research. The Appendices cover several side issues that, although not strictly part of the algorithms, are necessary to make them work.

Chapter Organisation

This chapter introduces the dissertation theme and its working environment. Starting from an external point of view, the RESOLV project is presented and its grounds discussed. In Section Two, the results from the RESOLV project are presented. Section Three steps into the hardware and software characteristics of the RESOLV system.

Section Four introduces the Localisation Module, which is the core of the dissertation. A brief overview of the state of the art is presented and the proposed solution is compared to known algorithms. Then, the problem specifications and operating conditions are stated as well as the performance goals to achieve.

Finally, Section Five characterises the data types used and commands used to relate the Localisation Module to the other elements of the RESOLV system.

1.1 Presentation of the RESOLV project

In recent years, interest on three dimensional (3D) scene representations grew larger, in particular outside the CAD community, the traditional users of 3D representation. The ever increasing computing power made 3D scene representation available to common home computers, broadening the range of potential users and applications, using different paradigms of world representation. In addition, the emergence of the Internet as a global communication medium soon required flexible, platform independent tools for 3D representation. At the cross-roads of new users and applications, increasing capabilities of robotics and computer hardware, new 3D representation paradigms and the Internet, the RESOLV project was established.

The RESOLV-**RE**construction using **S**canned **L**aser and **V**ideo project [LeeversD] was developed under the ACTS program of the European Union. The RESOLV consortium gathers research, education and industry partners from different European countries:

European Union Joint Research Centre (JRC), Italy.
University of Leeds (UoL), United Kingdom.
Instituto Superior Técnico (IST), Portugal.
Instituto de Sistemas e Computadores (INESC), Portugal.
RoboSoft, SA, France.
BICC Group, Plc. United Kingdom.
VERS Associates, United Kingdom.
Zentrum für Graphische Datenverarbeitung (ZGDV), Germany.

The RESOLV project's goal is to develop a general purpose tool for 3D acquisition and representation of indoor environments, operated through the Internet via a Web browser like interface. Opposite to the regular CAD use, where one goes from the plan to the building, the RESOLV system provides a symbolic 3D representation of spaces already built. Thus, it is a "reconstruction" tool. Also, it is restricted to structured or semi-structured environments and, in particular, it is optimised for indoor environments.

The RESOLV project envisages three broad worlds of applications:

1. The Architectural world: the reconstructed models should interest architects willing to fuse real buildings to their CAD models of new buildings, the construction industry where it may be useful to verify the compliance of the actual construction site to the plans and the archaeologists to swiftly create 3D models of the working fields as the work progresses.
2. The Virtual Reality world, which favours the maximum likelihood to the real world, requiring high quality texturing and lightening. Applications range from role-play

games to virtual museums or other heritage sites, digital scenery for cinema or television productions, teleconferences, multimedia interfaces where the user travels between scenes, and likewise, real estate industry.

3. The Simulation world: 3D scene representations may also be used for real time tele-operation of robotic devices in human-hostile areas, such as nuclear power plants or wrecked tunnels. This is the most demanding application because it requires a high degree of accuracy, also a high degree of likelihood and, most important, near-real time operation. Another related area of application is surveillance and training of personnel. This family of applications is at a very early stage of development and it will take longer to mature.

As a first step in the 3D-reconstruction area, it is very likely that the current RESOLV system satisfies only poorly to all of these end user requirements. However, the current RESOLV systems are likely to be the common root from which all dedicated solutions will span. The RESOLV solutions were shaped by the aforementioned requirements. To satisfy the architectural user, a high degree of accuracy - comparable to the precision used in construction industry - must be provided. To satisfy the amateur user looking for Virtual Reality impressions, an appealing scene representation, featured with an intuitive and user-friendly interface is necessary. Tele-operation adds the requirement of nearly instantaneous model update. The proposed solution meets all of these to a certain extent and its results are presented in a manner to be easily translated to other information coding paradigms.

Designing an efficient solution to reconstruct complex indoor environments calls for range data and video image data processing algorithms. Robotics engineering is required to design hardware of automated devices and program its associated software. Computer scientists are required to build the software architecture, the interfaces inside the sub-systems within the system and interfaces to the human operator. Designing for the Internet imposes open-system solutions, general-purpose hardware and bandwidth restrictions.

1.2 The RESOLV output models

The three-dimension models produced by RESOLV are a metric three-dimensional (3D) representation of a given environment as a set of elementary surfaces, or other related spatial features, such as vertexes or 3D blocks built from hierarchical sets of surfaces. However, such models are uncomfortable to the non-trained user, because the models' surfaces have no textures. To fill the extracted surfaces defined by the map features, an image is captured by a video camera or other robotic vision device. From this image, the textures are extracted and associated to each surface element on the 3D map

[SequeiraV_2], [NgKia_1]. The resulting textured model looks rather natural, such as the one presented in Figure 1. Unfortunately, the 3D nature of the models turns the presentations on paper into a poor experience. Only the interaction with an appropriate 3D rendering tool on computer delivers to the user the whole spatial perception of the represented environment.

The internal representation of models uses Virtual Reality Modelling Language (VRML), an open-platform format to view and manipulate 3D data through the Internet. This format may be easily translated to CAD formats such as DXF from AutoDesk AutoCAD. The associated textures are mapped in image bitmaps, coded in the JPEG format.
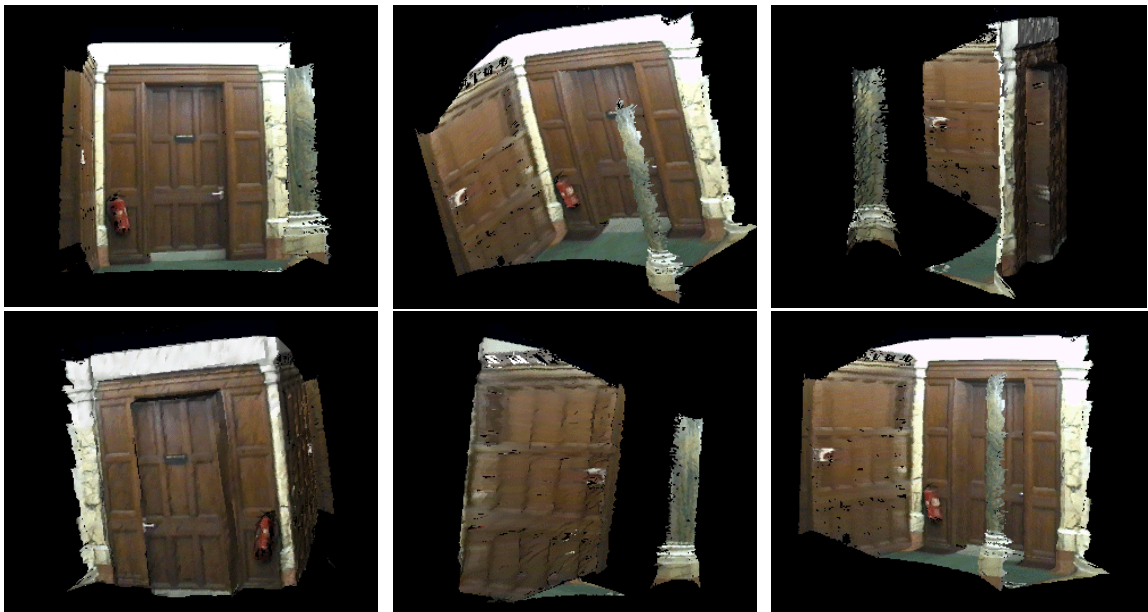

Figure 1 - Six images of a sample RESOLV model

The raw RESOLV models may be viewed instantaneously on an Internet browser. Currently, a VRML plug-in is often required to interpret the VRML format, which is not internally supported by many browsers. VRML readers allow not only the visualisation of the 3D model but also moving inside the model, turning the model around, hovering over it, inducing a strong feeling of *being there*.

Although the RESOLV project does not focus directly on advanced features of VRML and its extensions, some experiments were made in this direction. Adding video inserts, active points in the model to trigger events (such as "clicking" on a painting to see a larger image and the painting's description, or clicking on a door to move to a new room), mixing the model with streaming video, *etc.*, illustrates the versatility of the current solution and its capabilities for different types of applications.

1.3 The RESOLV system paradigm

The features required for the RESOLV project are: maximum 3D representation error less than 3 cm; 3D models described by planar and biquadratic surfaces based on a mesh of elementary triangular surfaces; realistic texturing; flexible acquisition platforms; automatic, real-time 3D model generation; remote operation through the Internet; friendly and intuitive user interface.

Several factors and constraints shaped the form of the solution proposed by the RESOLV consortium, which follows the specified requisites. The first and most important principle of operation is modularity, whose importance is emphasised by the fact that the consortium gathers groups of different competence, split around different European countries. Thus, it was most convenient to work on closed modules, linked by strict protocols. Derived from this first principle comes the second: it should be possible to add, replace or upgrade new hardware and software modules, without changing the functions that are not related to the changed parts. The third principle was to use standard hardware, models and formats as much as possible. The hardware is based on regular Personal Computers running MicroSoft Windows NT, data files are coded in text format, images are coded in common image formats, the 3D models follow the open standard VRML 2.0.

The RESOLV models are based on distance measurements and capturing images. To measure distances the Laser Range Scanner technology was chosen. This family of devices provides accurate range measurements, with a typical precision of 1 to 3 cm, on measurements ranging from 0.2 meter to 20 meter or more. In fact, technology has developed at a fast pace during the projects' course and the lasers used in 1995 are currently outdated by newer devices, with longer range of measurements, similar or better accuracy and faster operation, among other benefits.

The main drawbacks of the Laser Range Scanners are related to the nature of the target surfaces. In presence of poor reflectance surfaces, such as dark colours or soft materials (fabrics, furs, carpets, *etc.*), the laser beam sent to the target is returned with a feeble intensity due to absorption by the materials or multiple reflections. However, the most serious problems with lasers occur with high reflectance surfaces, such as mirrors and windows. A mirror changes the laser beam direction, producing an enlarged duplicate of the environment behind its surface. Glass windows react to the laser beam as transparent, partial mirrors, or perfect mirrors, depending on the glass type, thickness and angle of incidence. Some laser devices present additional security restrictions due to the beam power, which may be too strong for the human eyes, forcing people to leave the environment to be scanned and the operators to use protection goggles.

The images used to capture the surfaces' textures may be obtained by means of a video camera, using still frames or a digital camera. The first was chosen at the beginning of the project and, although newer digital cameras produce higher resolution images, the

current image quality suits the texture-filling task perfectly.

Once the main sensor technologies were chosen, the RESOLV system was developed around them. The main computer is the next element to add to the system. It is called the RESOLV PC or the Host PC. This computer drives the laser and video sensors and other auxiliary devices. The software running on the Host PC controls the laser and image acquisition process and performs the 3D reconstruction algorithms and the associated texturing algorithms. It requires also some modules related to robotics: a task scheduler to sequence operations, handle different work flows and overcome exception conditions; data management services and the localisation client, the object of the dissertation. The localisation module is required to associate the local data reference to the world reference. These three elements (laser, video and RESOLV computer) are the minimum elements to perform the assigned mission.

To build the 3D model of a structured environment one single point of view is not sufficient. On the contrary, many capture points are necessary to create a quality model. The device should screen surfaces occluded by other surfaces, approach some particular points to get a detailed view of the scene and also move from one area to another. Merging two partial models into a common one emphasises the importance of localisation in order to relate the different co-ordinate references.

To move the system between acquisition points, some type of cart is necessary. Moreover, to keep the system mobile, it would be convenient to embark a power source into it to be independent of external plugs and cables. These two additional elements are depicted in Figure 2. In case the cart is a mobile robot, other robotics system are required. A mobile robot should feature all the traditional robotics tasks, namely a trajectory planner to compute a safe path between data acquisition points; a navigation controller, which is responsible for driving the robot on its way and circumvent unexpected obstacles and handle exceptions [CastroJ], power controllers, additional motors and their associated drivers. In a mobile robot, Localisation co-operates with Navigation to optimise the robot movements and minimise collisions and dead-ends.
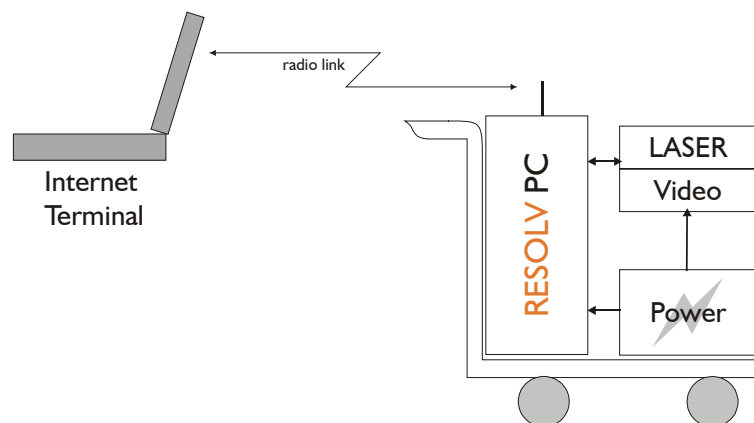

Figure 2 - The basic elements of the RESOLV system

Finally, since the human operator should be able to operate the system remotely through the Internet, a radio link between an Internet terminal or the networked Internet should be provided. The Internet terminal may be any type of computer linked to the Internet with a browser. It may be close to the RESOLV system or anywhere in the world. For practical reasons, however, in particular due to the limited bandwidth available, it is advisable to keep the terminal close to the RESOLV system.

This is a very simple sketch of the real system, denoting only the core modules of the RESOLV system. All the practical implementations of the system involve many other parts.

1.3.1   Experimental set ups

Initially, a manual cart was used to displace the sensing device around (Figure 3). Since the operator must push the cart manually between acquisition points, the supervisor terminal is mounted on top of the acquisition system and there is no need of a radio link. Batteries are installed on the bottom of the system to allow it to move away from power plugs. The two rear wheel axes pivot: they open to increase stability during operation and manoeuvre and close when tidying up the system. The sensor head features a Pan & Tilt motorised unit to turn the sensors around the scene. The sensor head height is fixed to the mast but may be changed manually by the operator.

This device was termed Environment Sensor for Telepresence (EST).

Soon it was realised that using an autonomous mobile platform was far more interesting, because it could be operated remotely and without local supervision (Figure 4). The new device included many traditional robotic features, such as odometry to measure the displacement of the acquisition system, a mobile tower to raise and lower the sensor head, ultrasonic sensors to avoid obstacles and a small camera to help the user to tele-operate the system. The radio link and the embarked batteries make it fully autonomous for a period of about six to eight hours before recharging is due. The mobile robot device offers many benefits: the odometers help to localise the system, the tower allows scanning under the tables and over higher surfaces, the remote control makes the operation more comfortable. Furthermore, the ultrasound ring provides an additional manner to perceive the surroundings, which could help solving the aforementioned glass and window problem although this feature is not currently implemented.

A new term was coined for the mobile system: Autonomous Environment Sensor for Telepresence (AEST).

Naturally, some other variants may be developed. For instance, the AEST may be operated locally as an EST, the EST may reduce to its simplest expression: a computer and the sensor head. Modularity was an important concern along the project; therefore, the various modules may be added, removed, changed or updated, with minor changes in the working environment. This is apparent in the two laser systems used in the EST and

AEST (Figure 3 and Figure 4, details in Figure 5 and Figure 6, respectively). The latter is more modern, therefore uses an enhanced device.

Figure 3 – EST, Environment
Sensor for Telepresence

Figure 4 – AEST, Autonomous
Sensor for Telepresence

The pursued modularity is very convenient to the RESOLV project, since this is only the first general approach to the study of 3D reconstruction of structured environments. Any future developments will tailor the current systems to their particular needs.
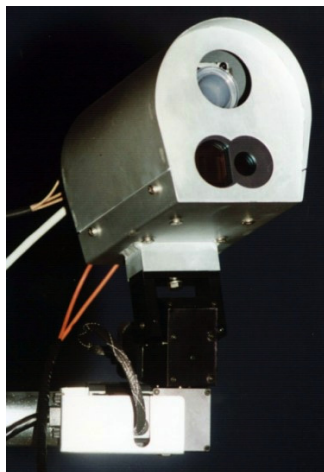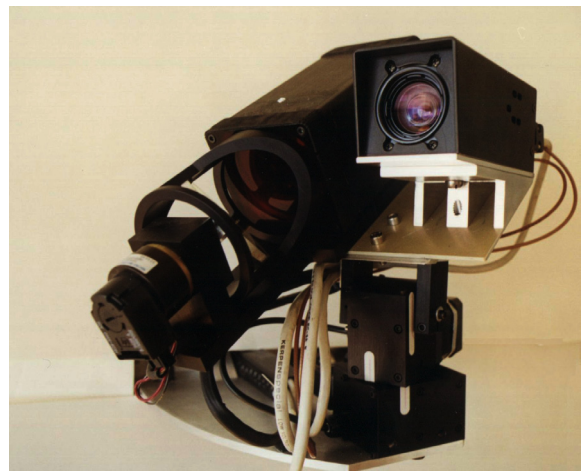

Figure 5 - EST Sensor Head

Figure 6 - AEST Sensor Head

In spite of the modularity and flexibility concern, the current systems do have some limitations. They are limited to indoor or regular ground outdoors and piecewise horizontal floor due to the lack of an inclinometer to measure the system angles relative

to the ground. In addition, they are limited on height (it is impossible to scan over very tall objects or under very low objects). Finally, the laser limitations on the nature of the target surfaces result in many gaps in the reconstructed models that must be filled later with computer graphics techniques and manual work.

Overcoming these difficulties is quite simple in some cases - for instance, a tripod could perform outdoors over uneven terrain - whereas others address fundamental limitations of the system. The laser reflectance limitations fall on the latter case and a different measurement method would be necessary. A new research line is to reduce the size and weight of the devices making them "wearable" by the human operator.

1.3.2   Hardware summary

The hardware used on the AEST and EST are somehow different although they share many elements. Since the AEST includes most of the elements it will be described in detail first (Figure 7), then the different elements in EST will follow.

Figure 7 - The main parts of AEST – Autonomous Environment Sensor for Telepresence

The core of the AEST is the RESOLV PC, which is a standard computer based on Intel architecture with a Pentium Pro microprocessor, 128 Mb RAM and several expansion cards, including a Matrox frame grabber to acquire the video images, a high speed interface to transfer the data from the laser to the computer and a card to control the vertical tower to lower and raise the acquisition sensor.

The RESOLV PC controls the Sensor Head (Figure 6) featuring a Directed Perception Pan & Tilt unit to rotate the acquisition cameras; a Acuity Research Laser Range Scanner, equipped with a rotating mirror to speed up data acquisition, named Line Scanner and a Sony Video Camera, producing PAL-standard images.

The RESOLV PC commands also the Tower, featuring a Galil motor, guided by side rails, to lower and raise the Sensor Head through transmission gears and strap. The RESOLV PC is also connected to a little Webcam from Connectix (currently from

Logitech), providing a navigation aid to the user tele-operating the system. The communications with the Mobile Platform are supported on two 38400 baud serial links, whereas the communication with the operator sit at the Internet Terminal is supported on a radio Ethernet link at 1.3 Mbps.

The Mobile Platform was manufactured by RoboSoft, SA, a robotics company member of the RESOLV consortium, based on their standard product Robuter II, with minor changes. It features two drive wheels and one free wheel, real-time control of the wheels, with odometry and velocity reading. It includes also a ring of 24 ultrasound sensors to measure the free space around the system and has good expansion capabilities to accommodate extra sensors or cards. The current implementation is using extra contact bumpers to protect the whole AEST system. At the core of the Mobile Platform, there is a Motorola 68040 microprocessor running at 25 Mhz, with 4Mb RAM. The computer is linked to the other elements via a VME bus card. Additionally, the Mobile Platform includes power converters to supply all other modules in the AEST system and two 24 V lead-acid batteries to supply the main 48V power source.

The Internet Terminal may be any type of computer provided it has a 1024x768 pixel colour screen and a fast Internet link.

The RESOLV PC, including the Sensor Head and the other devices are mounted in the white box, on top of the Mobile Platform. The package may be split in two parts for transport and the sensor head is detachable to be packed in a dedicated housing. The whole system is approximately 1.55 meter high, 0.7 m long and 0.5 m wide.

The EST uses the same RESOLV PC architecture. The main differences are in the Sensor Head featuring a Riegl Laser Range Finder, the absence of the Tower to automatically raise the head and the absence of the Mobile Platform. To replace the latter, a manual cart was developed, including the PC, the power converters required by the sensors and the batteries to power it all. The Internet Terminal was mounted on top of the cart and linked directly to the RESOLV PC. The EST is approximately 1.3 m high, 0.9 m long and 0.3 m wide, with its rear wheels hidden below the blue box.

Although many scenarios may be reconstructed by any of the two systems, one should take advantage of the AEST benefits (such as odometry, remote operation, vertical tower, *etc*.). However, under specific constraints, one should prefer EST as it is the case with stairs or in cluttered environments where operating a mobile robot may prove to be either too dangerous or too slow.

A more detailed analysis of the laser devices and the odometers, which are the sensors used for Localisation, is presented in Section 5.

1.3.3   Software architecture

The software architecture follows the same flexibility and modularity principles.

The full architecture includes three components: the Human Computer Interface running on the Internet Terminal, the Host Server and its associated clients running on the RESOLV PC (or Host PC) and the Mobile Robot Computer running a dedicated, real time operative system, Albatros (Figure 8) to perform the tasks associated to navigation.

Depending on the particular configuration, the Human Computer Interface and the Mobile Robot Computer may be absent. Only the RESOLV PC is required to build the models. Currently, the AEST is the only system using the three computers.
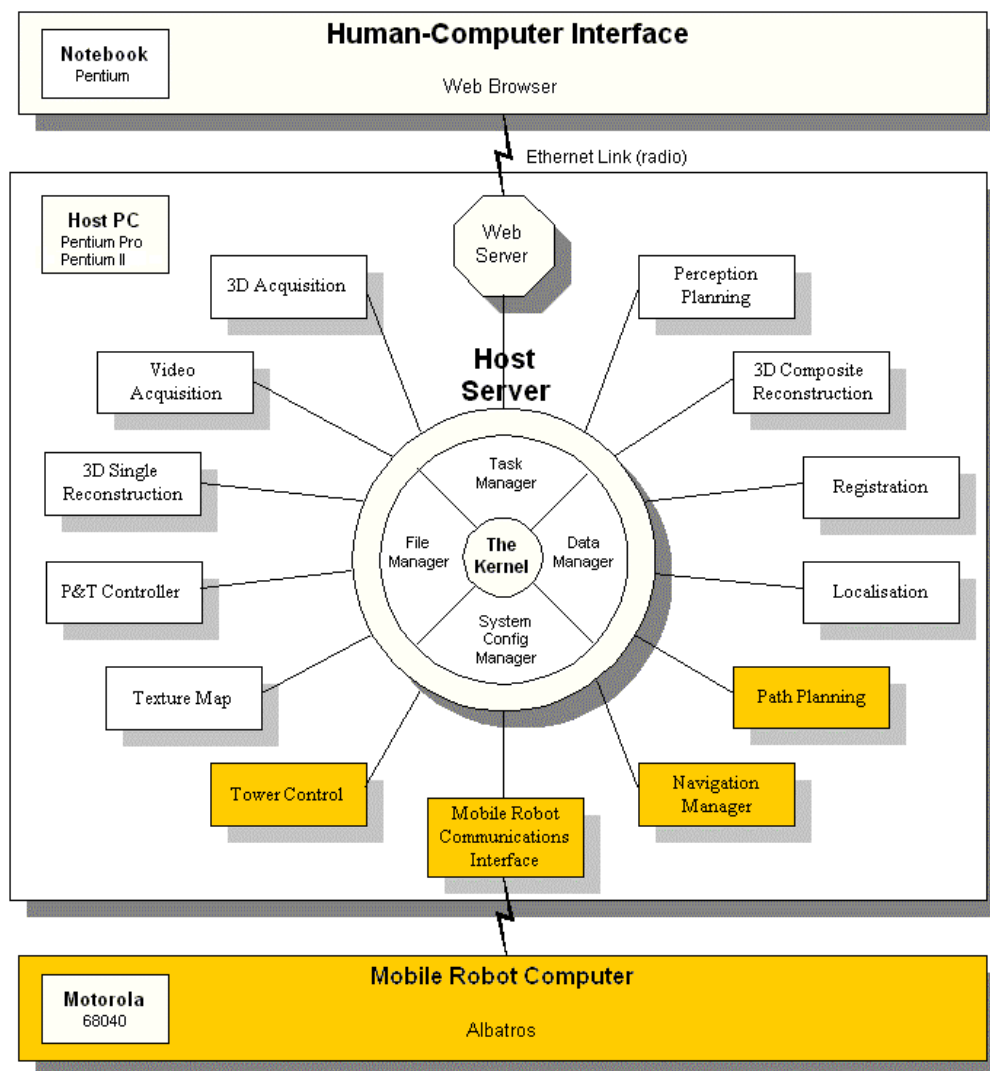

Figure 8 - Software Architecture

Starting with the Mobile Robot Computer, the Albatros is a multi-task, real time operating system, a commercial product developed by RoboSoft, SA, to be used with their platforms. It manages different concurrent tasks such as obstacle avoidance by means of the ultrasound sensors, collision detection with the bumpers, checking the distance travelled, and all other movement related tasks. The Mobile Robot Computer runs autonomously; it only requires the world reference and path to travel from the RESOLV

computer [CastroJ]. It may also operate under direct tele-operation from the RESOLV PC or the Internet Terminal.The RESOLV PC (or Host PC) runs the main software. The software is organised as a set of clients around a Host Server. The clients shaded in Figure 8 (Tower Control, Communications Interface, Navigation and Path Planning) are only run in the presence of the Mobile Computer, since their functions are specific to the AEST.

At the core of the program there is the Host Server developed by INESC [LopesFM], conducting and scheduling the operation of all the software clients. It contains four dedicated clients. The Data Manager is a common blackboard where the general real time data is posted. The System Configuration Manager loads the configuration files and sends the data to the clients upon request. The File Manager manages the intermediate data and final output files on the local hard disk and on the Internet server area as well (to be downloaded by the Internet Terminal). The Task Manager triggers the different steps of the reconstruction process and sequences all the client operations.

Each client performs a dedicated function, although some clients co-operate in order to fulfil a complex task. A brief summary is presented:

- Web Server: links the RESOLV PC to the Internet Terminal either through the radio link if the user is remote or through a cable if the user is close to the RESOLV PC.
- 3D Acquisition (JRC): controls the Laser Scanner and the associated Pan & Tilt unit (together with the P&T Controller). Upon request, it performs a two-dimensions or three dimensions scan. The results, a list of range distances, reflectance and other synchronisation data is made available to the client.
- Video Acquisition (UoL): controls the Video Camera and the associated Pan & Tilt unit (together with the P&T Controller). Upon request, it provides one grid of still images with PAL resolution (768x512 pixels, 24 bits).
- P&T Controller (JRC): the Pan & Tilt controller is the driver to control the mechanics of the Pan & Tilt Unit used to motorise the sensor head.
- 3D Single Reconstruction (JRC): uses the raw laser data to build a local model of the scenario scanned in one view.
- Registration (JRC): translates the Single Reconstruction model into global model co-ordinates and adds it to the pre-existent global 3D model.
- 3D Composite Reconstruction (JRC): merges the new elements with the previous ones, eliminating duplications and extending extracted features.
- Texture Map (UoL): associates the Video data taken from the images to the laser based 3D model.
- Perception Planning (IST): computes the next best capture point. This module may operate autonomously or follow instructions from the human operator.
- Localisation (IST): performs planar localisation based on laser data requested to 3D acquisition and, if available, odometry data supplied by the Mobile Robot. The localisation result is sent to the Data Manager as the new best posture estimate. This

module is the object of the dissertation.

- Path Planning (IST): computes the detailed path the AEST should follow to reach the desired next capture point.
- Navigation Manager (IST): monitors the AEST movement, checks for deadlocks and handles exceptions. It verifies also when the desired next capture point was reached.
- Mobile Robot Communication Interface (IST): links the RESOLV PC (or Host PC) to the Mobile Robot computer, updates posture data and ultrasound data (the Data Manager holds a copy of some of the most relevant Mobile Robot data for future extensions).
- Tower Control (IST): upon request moves the Sensor Head up and down, until the desired height is reached.

The Internet Terminal runs the Human Computer Interface, developed by INESC [GilP]. This is a Java Application, to control the acquisition process and the mission as a whole. A sample screen snapshot is presented in Figure 9.
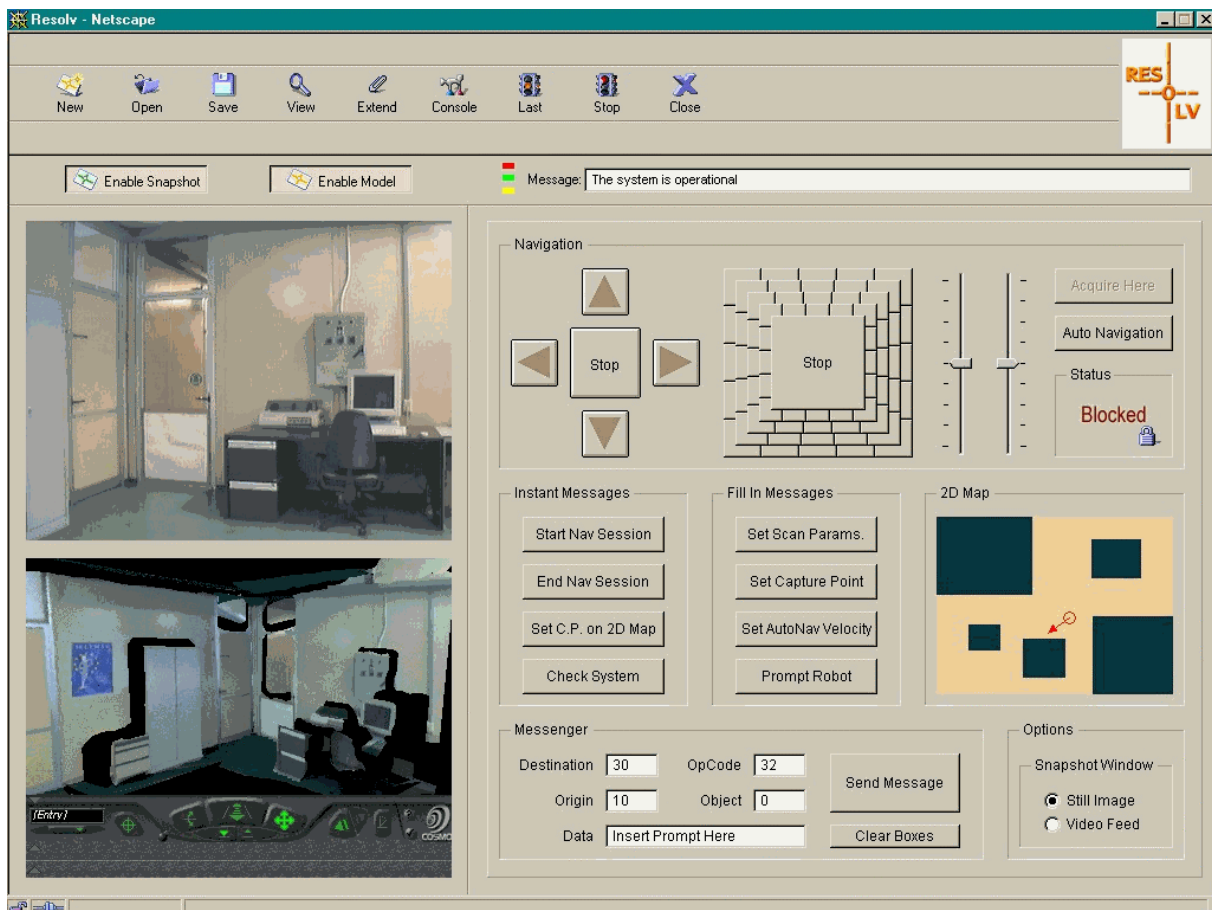

Figure 9 - The Human Computer Interface (HCI)

The toolbar contains the buttons to trigger the main actions, namely start a new model, save the open mode and stop the model creation.

The top left window shows a video snapshot of the scene to be reconstructed. This image may be supplied from either camera mounted on the system: the Video Camera used to capture the textures and the little Web Cam to help navigation. In the later case, it will be a real time video feed.

The bottom left window shows the model under construction, including the navigation buttons used to move within the 3D reconstructed world.

The large window on the right shows on top the commands to drive the system manually (tele-operation) and some status information. Just below there is a set of buttons to send messages to the RESOLV PC and a 2D map with a crude representation of the environment around the robot. This may be used by the operator to choose the next capture point. The last row contains an interface to access the Host Server and its associated clients using the internal syntax. This is reserved for expert users only. The Snapshot Window block on the lower right allows switching between video image sources.

Although the model post processing is not involved in the operation loop, some work has been done by the RESOLV consortium to refine the results and to obtain higher level information from the 3D models. Computer graphic techniques, such as filters to flatten the uneven surfaces, algorithms to fill in the gaps and correct different lightening conditions were used to improve the model quality. A second step is to compute volume information from the surfaces. An object extractor was created to this purpose: this module links adjacent surfaces to create polyhedral volumes that may be extracted from the model and manipulated on their own.

1.4 The Localisation Module

The Localisation Module design is subsidiary to the system set for the main task of RESOLV project. Therefore, it should require no extra hardware. When the project started, Localisation should be used on the Autonomous EST only. Shortly after, however, it was realised that Localisation could be valuable also with the EST, alleviating the operator from the uncomfortable and often inaccurate task of measuring distances and orientations with a tape. These design constraints narrowed the space of possible solutions since all methods based on odometry, ultra-sound readings, active beacons, etc., which are often available in autonomous robots, could not be considered. On the EST, the only sensors available are the Laser Range Scanner and the Video Camera on the Sensor head.

The nature of the RESOLV applications also constrains the type of solutions. Since the main task for RESOLV is to reconstruct a previously unknown environment, there is no *a priori* map, i.e., the map is created as the system explores its surroundings. In addition, artificial beacons can not be added to the scene, otherwise the system would not be autonomous. In some circumstances though, there is a 2D map of the environment,

usually a paper plan drawn by an architect. These maps encompass only the structure of the scene (walls, doors, windows), ignoring all the objects and elements inside it. When available, they should be used as a supplement or an external boundary for Localisation.

The Localisation Module was required to compute a planar posture estimate $(\mathbf{x}, \mathbf{y}, \mathbf{\theta})$ on horizontal floor, without any initial value, assuming the system is still. Precision requirements for Localisation were bounded by the 3D reconstruction algorithms at 0.1m position error and 17×10^{-3} rad (1°) orientation error. Notwithstanding, the error limits were reduced to 0.02m position error and 3.4×10^{-3}rad (0.2°) orientation error in trial environments, which proved to be quite challenging.

1.4.1   The state of the art

The Localisation solutions for RESOLV were tightly constrained by the problem characteristics. If it were possible to install temporary landmarks or to use extra sensor hardware, several common methods could be used. However, following the strict restrictions mentioned above the search for solutions already published is narrowed.

The solution presented in [DubrawskiA_1] requires laser range data only, too. The method postulates a relevant content of straight lines. It begins by the extraction of the lines in the range scan, which define the main directions in the scene. The set of directions is stored in a histogram. If a new scan is taken at a nearby position, the orientation change may be computed by cross-correlation of the direction histograms. The maximum of the cross-correlation indicates the shift on orientation of the robot with respect to the previous posture. The translation between scans is then computed from the shift on the parameters defining the straight lines.

The first part of the method, where the line parameters and directions are identified presents a resemblance to the Hough transform method, widely used in image processing. The line extraction and histogram computation may be computed in a parallel architecture, similar to a neural network, speeding up the results. The presented results proved very robust, provided the rooms had sufficient "linear" content and the displacements were small (approx. 0.2 m). Odometry data is present but is only used to verify the algorithm accuracy.

A class of solutions assumes the map is known beforehand. For instance, in [CrowleyJ], the known environment is acquired systematically by moving the laser sensor according to a grid pattern to create a thorough environment description. Instead of a real experiment, the laser scans can be acquired by a simulation based on a merge of a reduced number of real scans. Each individual measurement of a range profile with N samples is regarded as a dimension in a N-dimensional space. The Karhunen-Loève expansion is used to reduce the space dimension to a linear subspace with minimum

dimension. The elements of this space are termed "principal components of range data". In case there are no singularities, there is a reversible linear transform to express the posture space into the "principal components" space.

If the training set reflects all the relevant features in the environment, a scan taken on any point will be close to the elements of the training set, which can be extended by numerical interpolation. In practical experiments though, it occurs that more than one posture solution is acceptable and the posture must be tracked while the robot moves in the vicinity of the original point to confirm the correct candidate and discard the false ones. This process uses the Kalman-Bucy filter.

In [VestliS], a method closer to RESOLV's is proposed. The laser data is transformed into a set of lines or circle arcs by means of a geometric transform associated to a neural network. These features are combined to resemble natural landmarks, such as corners and cylinders. If the robot displacement is approximately known, a feature extracted from the range data should be close to the corresponding feature on the map. The match for each feature generates a possible space for the robot location. If multiple features are matched, the robot posture is known with high accuracy.

A related work [ArrasK] added the detection of artificial landmarks and an intense statistic analysis of the laser data. A hybrid position estimation method based on an Extended Kalman Filter encompassing extracted lines and artificial beacons is proposed. This work makes extensive use of the reflectance data to achieve the statistical classification of the range measurements.

These two approaches assume real-time continuous localisation, a high quality model of the operation environment and an odometer device to compute incremental movements. In both cases, the robots are assumed to operate in fairly uniform office scenes, which allows for strong constraints and statistical optimisations. This is not the case in RESOLV.

In [WeckesserP], a solution for 3D mapping is presented. This work is fairly related to RESOLV's goals, although the paradigm is different: it uses a 3 camera video system as an acquisition tool. Surface boundaries are its distinctive features and range data is translated into a similar description. To perform localisation the lines extracted from the range scan are compared to the lines in the world model and a Mahalanobis distance criterion is used to find a match. This solution complies with the constraints of the RESOLV system, except for the acquisition sensors. However, it should be interesting to try adapting it to RESOLV.

Another possibility arises from the use of some special "natural" landmarks: Global Positioning Systems (GPS). Even though the satellite constellation used for GPS is currently maintained by the United States of America Department of Defense, nearly every system on Earth can use it to accurately compute the position of a GPS receptor and

it is reasonable to expect that the system continues to operate in the following years. However, using GPS for in-door applications is still very expensive and the hardware and software required to achieve the accuracy stated above, using Differential GPS, make it non-practical for the RESOLV system.

Although the proposed algorithms are close to the mentioned approaches, specially [ArrasK], [VestliS] and [WeckesserP], it will be shown that the environment and sensor constraints can be relaxed. RESOLV's solution requires no initial posture estimate and odometry is a supplement only. It includes two complementary algorithms, enhancing its overall performance and minimising each algorithm's pitfalls. Last but not the least, it performs correctly even though only a small fraction of the surrounding scene is known while coping with textures with different reflectivities.

1.4.2   Localisation specifications and constraints

The Localisation Module depends on many other elements in the RESOLV system. The first class of elements to consider is the acquisition sensors. Several different sensors are available, depending on the RESOLV system. The AEST encompasses all sensors, namely: Laser Range Scanner, Video Camera, odometers and ultrasound sensors.

The proposed solution is based on the Laser Range Scanner for two main reasons. First, it provides distance measurements accurate enough for the proposed applications and its operation range is within the limits of ordinary indoor spaces. Second, the data are expressed in three dimensions in an easy, accurate and non-ambiguous form.

The Video Camera could be used either autonomously either as a complement to the Laser Range Scanner. However, it was not used for localisation purposes because the extraction of 3D information from a single camera is a challenging process and the Laser Range Scanner yielded high quality data, fit for 3D analysis.

The odometers, which are present in the AEST only, are used as an initial posture estimate to refine the localisation process. The odometry accuracy depends largely on the nature of the ground surface and its errors accumulate as the mobile robot travels.

The ultrasound sensors are not used since they have a very low angular resolution, a shorter operating range and are very sensitive to small obstacles close to the ground which may be not present in the map. However, they could provide valuable help in the presence of mirrors and widows.

A second context element is the environment map to which the posture estimates computed by Localisation will refer. The goal of the RESOLV project is to create a 3D map of the reconstructed scene. Quite often then, there will be no other map than the one created by the 3D reconstruction process.

This map includes all surfaces already reconstructed (Figure 10). To increase the operation speed, Localisation is based on 2D data only. Hence, the 3D surface map is cut

by a horizontal plane at the Laser Scanner height, which will be compared to a 2D circular scan on the horizontal plane.
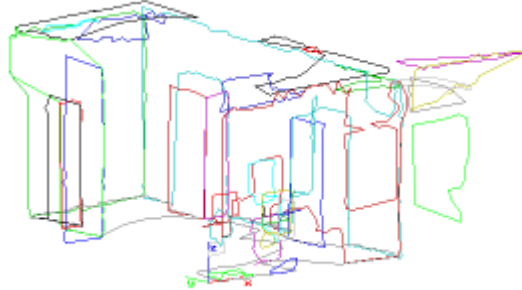

Figure 10 – A sample surface map

During the project trials it became apparent that the use of a pre-existent map, even in case it describes the environment only poorly, would be very useful to overcome some laser limitations (such as the aforementioned window problem) as long as localisation could manage to "read" it. If it were possible to use the *a priori* map only, then the 3D reconstruction process, which is time consuming, could be postponed to a later batch procedure, reducing the total acquisition time. In case both maps are present, it should be possible to analyse them in a coherent reference, with two benefits: when the acquisition session starts there is already a map of the environment; when the laser fails to extract a given feature, localisation can rely on the *a priori* map to fill in the gaps. The extension of Localisation to the *a priori* map has been implemented with some restrictions on the simultaneous use of two maps, where problems arose when inconsistencies between the two maps had to be solved.

The reconstructed map is created with Laser Range data and Localisation will rely on Laser Data, as well. The akin nature data used on both processes emphasises some pitfalls of the Laser Range data. If the reconstruction process fails to identify correctly a given feature on the environment – a mirror, for instance – it is very likely that the error is repeated with Localisation. Under these circumstances, a different sensor would be valuable. Currently, the laser data is analysed with dedicated statistics to detect irregular behaviours but no other sensor is used to solve this problem entirely.

The Localisation Module is also closely tightened to the other software clients running in the RESOLV system. The interface between Localisation and the remaining modules will be discussed in the following section.

1.4.3   Different solutions to different problems

From the referenced literature, it was apparent that no single solution could solve the proposed problem, in case the map is unknown beforehand. First, there is no algorithm to compute the absolute localisation of the robotic device without any prior

posture estimate or *a priori* position and orientation postulates.

Although incremental methods are available, these would put heavy constraints on the displacement of the device. Under certain circumstances, one can accept the robot travels by small steps and stops often to re-localise itself. However, if the scene to be reconstructed is quite large (more than 20 meters), it will take too long to travel step by step. When tele-operating the AEST it would be also necessary to persuade the human operator to stop at very often to localise the system.

A second problem arises if the environment is filled with objects. Since it is necessary to go behind the objects, the previously scanned areas would be occluded. Moreover, when scanning cluttered environments the system leaves the reconstructed areas behind its "back" while looking forward to new, unmapped areas. Therefore, the localisation method should perform correctly even if only a small fraction of the reconstructed environment is visible.

Briefly, the Localisation algorithm should require no correlation between successive postures and it should require a minimal set of features common to the map and the laser scan at each acquisition point. The algorithms proposed in this dissertation include such a method, based on feature extraction and matching. They require only a map of the environment (either from 3D reconstruction or an off-line map) and one single laser scan. However, its accuracy was considered unsatisfactory. Moreover, since in many operating conditions there is additional data available and also a greater correlation between the laser scan and the environment map, it would be foolish not to enhance the proposed solution.

Hence, the proposed Localisation Module is organised as presented in Figure 11. If no *a priori* posture estimate is available, the Frame Localisation algorithm is called. This uses only the laser scan and the available map (usually from 3D reconstruction but it can use an external map or in some cases both maps merged). The results are passed through a Likelihood Test, LT, and the successful candidates are a coarse estimate of the current posture (in office rooms used on trials, the expected accuracy is 0.05m, 17.4×10^{-3} rad or 1º). The results of the first stage are fed into the second stage, the Approximate Localisation algorithm, which is based on correlation analysis of the laser and map data, followed by co-ordinate transforms. The final result is submitted to the Likelihood Test, now with constrained parameters (typical accuracy: 0.02m, 3.5×10^{-3} rad or 0.2º) and, in case of success, the final estimate is issued on the common blackboard. In case there is a first posture estimate, supplied by odometry or by other external source the first stage is bypassed. Under some assumptions, discussed in Chapter 4, the Approximate Localisation can bypass an external map, building its own map from previous laser range data.

To increase robustness a closed loop solution was implemented. This will be discussed in Chapter 5.
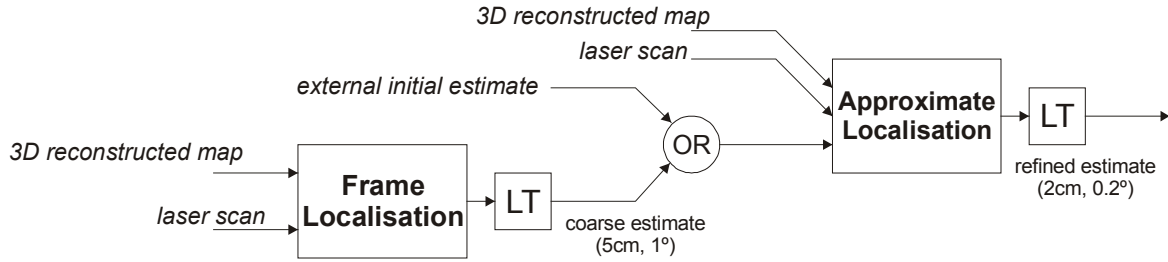
Figure 11 – The sequence of algorithms

The proposed solution meets the precision requirements and minimises the computation load while optimising the data acquisition process. It should be noticed that this approach is adequate for the EST (the manual trolley) and the AEST (the mobile robot). For instance, in the EST there is no odometry, but a coarse estimate can be computed by Frame Localisation. If Frame Localisation does not provide a posture estimate (because it fails or there is no suitable map) it is possible to input an external initial estimate (measured with a tape or any other method) and start the second stage to achieve a refined estimate. In the AEST the same method may be used with the extra help of odometry, used to track the robot movement between acquisition points.

1.5 The Localisation Interface

In addition to the final 3D models with texture, the RESOLV system produces a wealth of information to share among the software clients working on the RESOLV PC. Although the intermediate files are seldom available to the operator, they are the core of the information processing and may be accessed after the data acquisition.

This procedure allows separating the data acquisition from the data processing, which is often useful due to the limited time on site granted for operation. Moreover, this procedure also allows the development and analysis of each individual module, stressing the importance of the modularity principle already stated.

Since this dissertation documents on the Localisation Client only and the RESOLV project has been introduced in the previous sections, the focus will now concentrate on the Localisation Client issues. The Localisation Client interface definition encompasses the related input and output data, and the messages and commands exchanged between the Localisation Client and the other modules of the RESOLV system. In the future, different solutions for Localisation can be implemented, provided they comply with the following specifications.

1.5.1   Localisation client model

The Localisation Client (abbreviated "Localisation") is an application running under

the event/response model. The application is idle until a message or set of messages coming from other applications (also named "clients") brings a request to Localisation or a reply to a Localisation request. Depending on the incoming message the adequate procedures should be carried out and in the end, a message sent to inform about the new condition or answer to the client request.

Two methods are used to share information between clients: short amounts of data are exchanged in the bodies of the messages while large amounts of data are exchanged through files and the name of the file is passed in the message body.

According to the software architecture running on the RESOLV PC, the Localisation Client communicates directly with the following clients (Figure 8): Data Manager, System Config Manager, Task Manager and 3D Acquisition. It also uses data supplied by 3D Composite Reconstruction and File Manager.

Messages are handled in sequence of arrival, using the queue system provided by the operative system. There is no priority or semaphore mechanism to select between messages in queue. Therefore, the Localisation Client should be programmed to handle messages as fast as possible, to process swiftly the incoming messages waiting in the queue. If there is any long procedures to perform, they should be launched concurrently with the message handle mechanism.

The message syntax is strictly defined, including data formats and request permissions, *i.e.*, which clients are allowed to issue which messages. However, the model of message definition was kept open to allow for new message definitions and extension of current messages. The message format is as follows:

Origin	Destination	Operation Code	Object Code	Data
1 byte	1 byte	1 byte	1 byte	*n* bytes

Figure 12 – Message format

The reconstruction process is organised in iterations. At the beginning of each iteration, Localisation runs to update the posture estimate. Afterwards the remaining modules are invoked to perform the 3D reconstruction and navigation tasks.

The message sequence of a sample localisation update is presented in Figure 13, where the details of the particular implementation of the Localisation are hidden and the Clients exchanging messages with Localisation are emphasised.

The Task Manager triggers Localisation at the beginning of each iteration. Localisation requests the session data (*e.g.*, iteration number and time-stamps) to Data Manager and then requests a line scan to 3D Acquisition. The Line Scan is a particular type of laser scan: it consists of a circular sweep of the laser beam at constant height. The laser data, expressed in 2D polar co-ordinates is swiftly acquired. Meanwhile, Localisation loads the reference map, from 3D Composite Reconstruction, an external source or an internal map computed from previous iterations, depending on the current operation specifications.
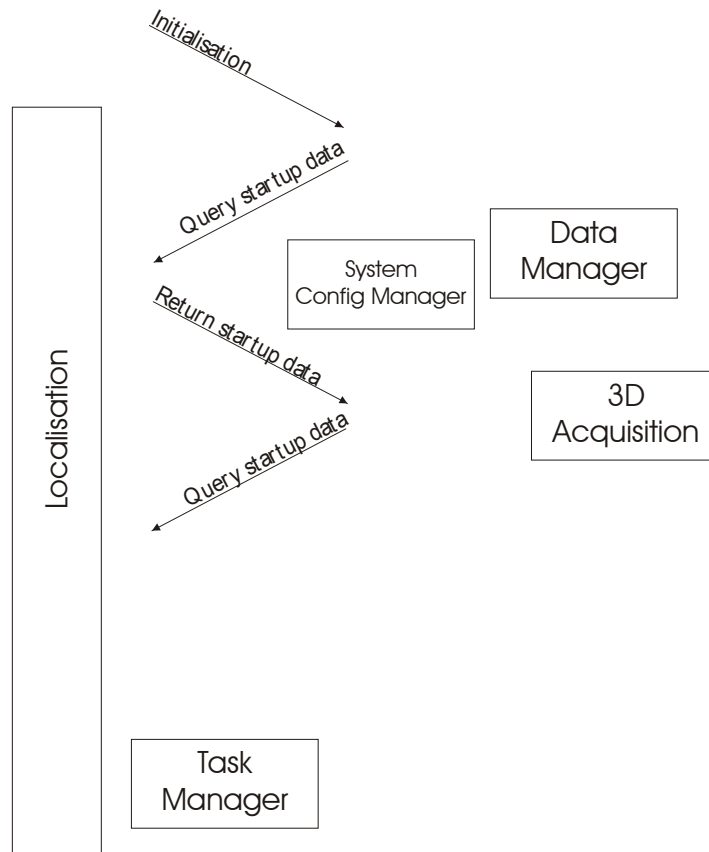
Figure 13 – The message sequence for Localisation

When the laser data is handed to Localisation, the most recent posture estimate is requested to Data Manager to decide which localisation algorithm to run. When this last data is returned Localisation runs internally to compute the updated posture estimate. Upon completion, the new posture estimate with associated uncertainty is issued to Data Manager, the Task Manager is informed of task completion and whether it was successful. Localisation is then idle until the next iteration.

1.5.2 Laser data

The Laser Range Scanner is the main sensor for localisation purposes since it provides the distance information used to compute the spatial features around the system.

The typical Laser Range Scanner (Figure 14) features a laser beam which measures the time lapse a burst of energy takes to travel from the scanner to the obstacle ahead and return to the detector sensor on the scanner. The time of flight is divided by two and then by the speed of light on vacuum, returning the distance from the scanner to the obstacle. In order to transform a punctual information into a 3D depth map some sort of rotating device is used.

One solution is a two degree-of-freedom Pan & Tilt Unit like the one in the EST (Figure 5); the detached laser range finder is shown in Figure 14. The Pan & Tilt Unit

sweeps the laser beam in azimuth and elevation, between predefined boundaries yielding a 3D depth map within a solid angle in front of the scanner. In fact, some of these devices are termed Laser Range Finders and only when they are equipped with a sweep mechanism are they called "scanners". Some newer devices, like the laser head on the AEST (Figure 6) have their own rotating device to scan the environment. This head - presented detached in Figure 15 - features a 1D rotating mirror to sweep a vertical disc perpendicular to the laser beam (assuming the camera is horizontal). To obtain a solid angle, the Pan & Tilt is still required, but only the azimuth angle is used. Other heads may have more complex mirror arrangement or an internal Pan & Tilt Unit and therefore external Pan & Tilt units are no longer required. As a rule of thumb, the smaller and lighter the moving part is, the faster the scanner will sweep the scene.

Figure 14 – Simple Laser Range Finder

Figure 15 - Laser Range Scanner with 1D spinning mirror

The Laser Range Scanner data usually comes in two forms: range and reflectance. The latter is only an auxiliary source of information, similar to a grey scale image of the scene. The two types of images are shown in Figure 16 and Figure 17. Each figure shows a 386x425 pixels image corresponding to a single scan taken with the scanner depicted in Figure 15 [SequeiraV_1].

Figure 16 - Range data

Figure 17 - Reflectance data

The edges are rounded due to the geometry of the scanning device, leading to an even angular grid, which aggregates the data in the centre of field of view and disperses the data on the image boundaries. Figure 16 presents the range data, where lighter grey means closer range and darker grey means farther range; white may also mean no valid data. The reflectance image depicts the fraction of incident light returning to the laser device. Since the lasers used are close to the visible wavelength this is close to a grey scale image taken with a standard camera.

From the range data, it is possible to compute the 3D profile of the scene, whereas from the reflectance data on can compute the reliability of the range data for each point. If the returned energy is feeble (reflectance is dark), the degree of confidence on the range measure should be reduced. Reflectance may also be used to detect more refined features: because the quantity of the energy returned depends on the angle of incidence of the beam on the surface, a sudden change in reflectance means either a change of material or a corner. Moreover, if one compares the reflectance image to a video image it should be possible to associate features on the reflectance image to similar features on the video image. Then, it is possible to associate the range information to the video, through the reflectance image, due to the one-to-one correspondence between the pixels in the reflectance image and pixels in the range image [NgKia_2].

The operation range depends on the technology used and the quality of the device. The effective range of the lasers used with RESOLV varies from between 10 to 20 meter. More recently, newer scanners were introduced with an operational range of 200 m.

The statistics of Laser Range data is quite complex, especially because it depends to a large degree on the statistics of the surrounding environment, which is often unpredictable. The typical accuracy lies within a 1 cm to 4 cm range. Some devices have constant accuracy throughout their entire operation range (an "error cylinder"), while others have "cones of error", i.e., varying standard deviation depending on the range [SequeiraV_1]. This is the main source of measure incertitude and it is characteristic to the Laser device. A second important source of incertitude results from the angle of incidence. As a first approximation, the quantity of energy returned depends on the angle of incidence, β; it is proportional to the **cos(β)**. The third source of incertitude is, as mentioned before, the environment characteristics. Since RESOLV is a reconstruction tool, it should be prepared to cope with any type of material present in human environments. The information about the materials is concentrated on the reflectance image, in a poor manner, however.

Finally, an odd form of incertitude results in the presence of mirrors and windows. This is a fundamental problem with the Laser and little can be done about it. It is possible to detect some windows from the reflectance data but a new sensor, with a new measurement paradigm is required to effectively cope with the glass and window problems.

1.5.3   Odometry data

When using a mobile robot odometry is used as an accessory source of localisation information. Assuming the robot is moving on flat floors, and in the absence of additional sensors, odometry provides 2D position and orientation co-ordinates.

Opposite to the laser that provides absolute measurements, odometers count the wheel revolutions, hence they produce relative information and the incertitude associated to the estimates increases as the mobile robot travels. Therefore, the odometry counters should be reset often enough to keep the incremental error within reasonable bounds.

Since odometry is available only in some RESOLV configurations, the Localisation Module uses odometry only as a supplement. It may be used as a start value to initiate the algorithms, to verify or benchmark the Localisation algorithms or as a backup posture estimate if, for any reason, Localisation fails.

The incertitude associated to odometry varies from less than 1 cm if the mobile robot is operating on hard ground and the odometry estimate is reset at after each short move, to a large value, if the robot slips on the ground, the odometry estimate is not reset or the distance travelled between stops is large [CastroJ].