On building 3D maps using a Range Camera: Applications to Rescue Robotics

Raymond Sheh, M. Waleed Kadous and Claude Sammut
ARC Centre of Excellence for Autonomous Systems
School of Computer Science and Engineering
The University of New South Wales
Sydney, NSW, 2052, Australia
[rsheh|waleed|claude]@cse.unsw.edu.au

UNSW-CSE-TR-0609
April 2006

THE UNIVERSITY OF
NEW SOUTH WALES

SYDNEY · AUSTRALIA

Abstract

It is critical in many mobile robotics applications to characterise the presence and position of objects around the robot. This is the case whether the mobile robot is under autonomous or teleoperative control.

In this paper, we examine the use of a CSEM SwissRanger SR-2 3D range camera which allows the generation of dense, accurate 3D point clouds around a mobile robot. Combined with other data sources, such as video cameras, this allows the creation of 3D maps that can be used for "fly throughs". Furthermore, this same technique allows a teleoperator to very accurately place landmarks within the 3D maps.

As this device is still somewhat prototypical, we also discuss some of the issues associated with the use of this device. The test application was the 2005 RoboCup Rescue Robot League, a competition that simulates robot-assisted Urban Search and Rescue (USAR) tasks and places great importance on effectively generating maps.

Novel techniques for processing the raw measurements from the sensor, and its use to create maps of mock disaster sites are discussed. The maps generated, part of Team CASualty's entry, were received very well by the judges of the competition and were unique in their combination of 3D, colour and thermal information, and the automated way in which the placement of landmarks and other annotations were performed. The maps were instrumental in the team's achievement of 3rd place.

1 Introduction

Mobile robots that must deal with the real world almost always need to be able to sense the presence of objects in their environment. Some also need to record this presence and/or locate themselves relative to these objects – to generate a map and/or localise. Many sensors currently exist that can be used to determine the distance to objects and, with appropriate algorithms, generate a map of an environment. Examples that have been used with varying degrees of success include stereo vision, scanning time-of-flight lasers, and infrared and ultrasonic rangefinders.

However, until very recently, no commercially available sensor was available that could produce a dense 3D range image in "one shot" and very few could produce 3D range images at all. The CSEM SwissRanger SR-2 range imager is one of a new family of sensors that generate real range images at video framerates. In this context, a range image is an image where the value of each pixel represents the distance from the sensor to the nearest object along the line of sight of that pixel.

This paper presents results using this range imager [5] for the purpose of generating dense, textured, automatically annotated 3D maps of an unstructured environment. The generation of such maps using traditional sensors has been very difficult in the past. A dense map consists of many points, not just edges or corners, and can be used to detect convex, concave or missing surfaces as well as small objects in the environment. When combined with other sensors, such as web cameras, textured 3D models can be constructed allowing the operator to "fly through" the scene.

1.1 The RoboCup Rescue Robot League

The sensor and mapping system was tested on the Robocup Rescue Robot League (RRL). The aim of the RRL is to deploy a robot in a mock collapsed building, find simulated victims, identify their signs of life and generate a map, annotated with the locations of victims and significant landmarks. A human rescuer should be able to follow this map in order to find and rescue the victims. The map is scored based on how closely it resembles the actual layout and how accurate the marked positions of victims are relative to easily identifiable landmarks. Operators have no prior knowledge of the environment layout and must operate the robot from a remote location. The operator has a limited time in which they can complete the mission. The simulated victims have signs of life that range from movement and skin colour to heat emissions and sound.

The environment is highly unstructured and features of interest are rarely vertical so traditional 2D maps generated by horizontal scan laser rangefinders are of limited usefulness. In particular, mapping systems aboard these robots cannot assume that the floor on which they rest is level. Despite being unstructured, the environment is stable and reproducible so comparisons may be made between alternative mapping systems. As the RRL is a competition with a significant mapping component, there was ample opportunity for comparison. Finally, being a mobile robot league, implementation for this domain forced the system to be deployed "real time", aboard a robot that was small and mobile enough to overcome the unstructured terrain that it was to map.

To generate maps with useful additional information, for the purpose of the

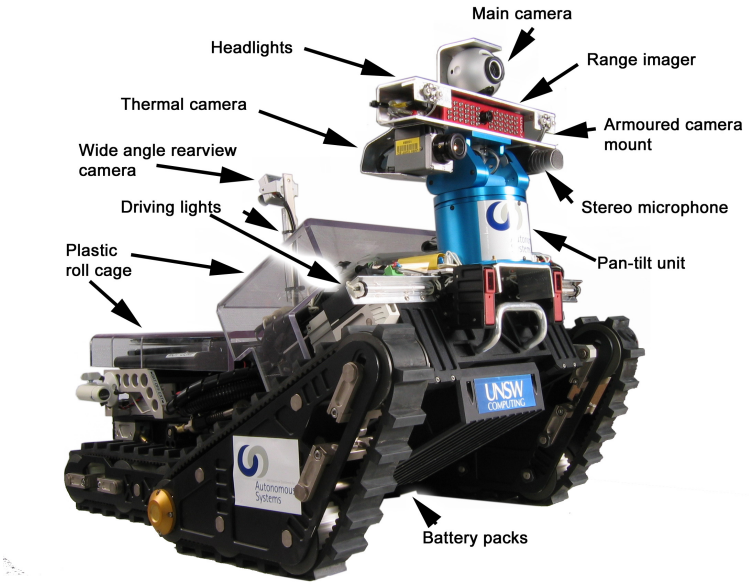

Figure 1: Oblique view of CASTER

competition the range imager was augmented with two additional sensors. A Logitech QuickCam Pro 4000 colour camera was used to provide colour information. This colour camera was also the main camera that was used by the operator for driving and locating victims and landmarks. Thus the operator could accurately locate objects in 3D space relative to the robot very easily. A FLIR ThermoVision A10 thermal camera was also co-located with the range imager and provided the temperature of observed objects. This was critical as the victims sought by the robot emitted heat. These sensors were deployed aboard the USAR Research Robot CASTER, based on the Yujin Robotics ROBHAZ-DT3 [8]. The robot and sensors are shown in Figure 1.

1.2 Objectives

The main aim of the mapping subsystem aboard the rescue robot CASTER was to generate dense, textured 3D maps of its environment. Ideally, the gathering of the necessary data should be done "on-the-fly" but failing that, stop-and-scan approaches could be considered. Secondary aims were:

- To automatically locate landmarks and victims in 3D space relative to the robot with minimal user input
- To automatically register consecutive scans into a global map
- To be able to gather necessary data despite the robot not being level
- To place a minimal size and weight burden on the robot
- To place a minimal time and attention burden on the operator

2 Prior Approaches

Until recently, maps generated by robots have tended to be two dimensional. Flat laser scanners generate maps that work well for environments such as offices where primary features are vertical, such as walls and doors. However, in many real-world environments there are no reliably vertical surfaces. Perhaps more importantly, features of interest are rarely vertical, instead important features include debris on the ground, holes, overhangs, low clearances and other factors that are only visible on a 3D map.

Various groups have produced solutions to the generation of 3D maps. Thrun et al [7] have done trials with dual line scanning laser rangefinders, one mounted horizontally and performing conventional 2D simultaneous localisation and mapping (SLAM) while a second is mounted vertically to provide wall profiles. Whilst successful in mapping indoor environments and, in some tests, an underground mine environment, this approach suffers from an inability to deal with tilt and roll of the scanning platform, especially if this changes quickly as happens when traversing unstructured terrain. The use, in some cases, of a flat wall and flat roof assumption also causes problems in such environments.

3D mapping first appeared in the RRL in 2004 when Kurt3D from AIS Fraunhofer [6] presented a solution based on a SICK LMS-200 laser rangefinder, mounted on a pivot such that the scan plane could be rotated perpendicular to the usual scan axis. By slowly rotating the scan plane a spherical range map could be generated. Whilst effective, this solution has several drawbacks. The LMS-200 laser rangefinder is extremely heavy, powerhungry and somewhat large, precluding its use on small to medium sized high mobility robots. The robot needs to be stationary while the scan is being produced, as does the scene. Determining the location of any given point in the camera image would have required a complete scan (or at least repositioning of the laser scanner).

The appearance of new, significantly smaller laser range scanners, such as the Hokuyo URG series [2] alleviate some of these issues. However their slower scanning speed results in a significant tradeoff – at 10 scans per second a 160 line range image takes 16 seconds. Despite this, the much larger horizontal field of view and light weight of this makes it deserving of further investigation.

One area where 3D mapping using laser range scanners has already been performed with great success is in the capture of architecture. Architectural laser scanners from companies like Riegl and Cyrax have been used to capture high resolution 3D data of the interiors and exteriors of buildings, factories and other large structures. This has then been combined with high resolution colour images to form dense, textured 3D models [4]. However, their suitability for small mobile robots in unstructured environments is limited. These devices are similar to the rotating laser scanner method above and so only work with a static scene – 24 seconds per scan is considered high speed. The devices also tend to be extremely heavy, beyond 10kg. Many of these systems rely on the presence of known features, such as straight lines, in the scene, an assumption that cannot be made in a general, unstructured environment.

Surprisingly, map generation is not one of the prime applications of the SwissRanger. Instead, current applications are in sensing of objects, such as humans, cars or factory goods in a static environment. Examples include occupant detection and characterisation in automobiles, intelligent door, gate and elevator controllers, biometrics, security, machine vision for quality control and the arts

[1].

Prior to the wide availability of range imagers, stereo vision was often used to generate range images. By measuring the disparity between corresponding points in two images taken from slightly different viewpoints, the distance to those points in 3D space can be calculated, in a similar way to the way human eyes determine distance (independent of context information). However, determining which points correspond in the two images is difficult, especially for scenes that either have few visual features, such as blank or dust-covered walls, horizontal lines or objects that have certain repeating patterns. Accurate disparities and therefore range measurements are generally only available for points corresponding to visual edges; dense range images are rarely possible in real world environments from stereo vision.

Unless all areas of the environments being mapped are visible from a single point, multiple measurements from such devices need to be combined to form a global map. Generally these methods fall into two groups. The first group attempt to find some sort of structure in the point clouds that result from these measurements, such as walls and floors [7]. While such methods work well for structured environments – architecture, offices, factories – they fail when confronted with features that they cannot model or environments with little structure, such as a rubble pile.

The second group make use of the points themselves, perhaps with some additional processing and perform registration of adjacent measurements with few assumptions on the underlying structure. Most of these methods stem from the Iterative Closest Point [6] algorithm and perform some form of gradient descent on a cost function. Additions to improve the finding of correspondences, such as spin images [3] have also been developed. These methods can be very impressive in their ability to unambiguously match a point cloud with an existing model. Unfortunately, they are computationally very expensive due to the inability to abstract points into a small number of features.

3 Characteristics of the Range Imager

We used a CSEM SwissRanger SR-2 for these experiments. This device uses 48 infrared LEDs to provide near-infrared illumination of the scene, modulated at 20MHz. Each pixel in the image sensor samples at 80MHz, allowing the phase, amplitude and offset of the returning 20MHz signal to be determined. By measuring the phase shift, the distance of the reflecting object can be determined so long as it is no further away than 7.5m [5].

This sensor has a variety of characteristics that make it suitable for 3D mapping. As it does not need to scan the environment, the time taken to obtain a range image is instantaneous – in fact this device can provide a 160x124 pixel range image at 30fps. Combined with a field of view that is very similar to that of a normal camera, this allows the sensor's data to be easily integrated into existing systems that deal with images, such as robot driving displays.

The range images obtained are stable and measurements do not vary significantly as target materials or other objects in the scene change. Indoor lighting was also found to have a minimal effect on the range images – in practice only pixels that directly observed lights or direct reflections of lights were overloaded. In addition to a range image, the device also provides near-infrared images of

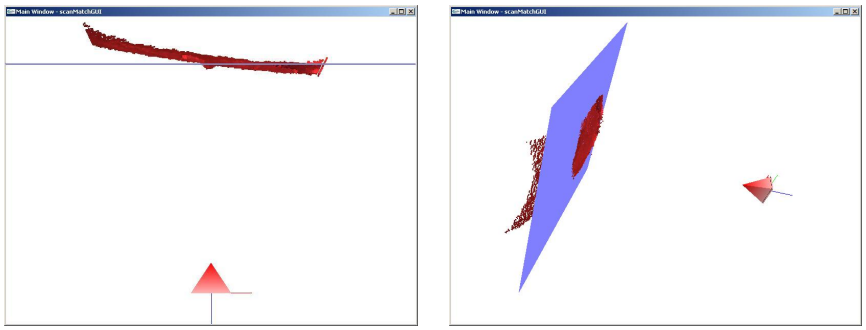

Figure 2: Projected measurements (red) of a flat wall at a distance of 1.5m assuming a uniform sensor. Left is a top-down view, right is an oblique view. Blue line (left) and plane (right) represent reference points exactly 1.5m from the sensor (red arrow).

ambient intensity and reflectance. These additional images allow overloaded pixels and pixels with too low a return signal to be detected and filtered out.

While this device does have a speed advantage over solutions that involve tilting and/or scanning laser rangefinders, it is somewhat lacking in image characteristics. The field of view of the sensor is limited to 45° so some scanning is required for full coverage of the environment. The lens of the range imager has very low geometric distortion – in practice barrel distortion correction was not required. However, a low depth-of-field results in significant focal blurring in the depth image causing bridges in depth discontinuities, rounded corners and other lost details. Focusing at approximately 3m was found to be a good compromise.

The use of phase to determine distance introduces another severe problem that limits the usefulness of this sensor – distance aliasing. Objects beyond 7.5m are reported as being somewhere in the region of 0-7.5m. A very obvious example of such a situation is when observing a long corridor. Sections within 7.5m of the camera appear straight but sections beyond 7.5m appear to "curl inwards", back towards the camera.

Finally the sensor itself has considerable variability across its sensing area. Figure 2 shows the projected measurements from the SwissRanger, assuming a uniform sensor. Distortion in the measurements is very apparent, especially towards the corners. In order to correct this distortion, a per-pixel calibration was performed. The sensor was pointed at flat surfaces of varying known distances and linear regression was performed on the resulting measurements. A lookup table with two entries per pixel – offset and multiplier – was created to convert the raw sensor measurement to a distance measurement. Figure 3 shows the result of this per-pixel calibration. Some distortion is still evident as the mapping from sensor measurement to actual distance is slightly non-linear but this accuracy was regarded as sufficient for mapping. There were also three non-compliant ("stuck") pixels in our sensor that were filtered out of this data.

The other two image sensors provided each point in 3D space recorded by the range imager with colour and temperature parameters. Due to their close proximity to the range imager and the similar fields of view of the three sensors,

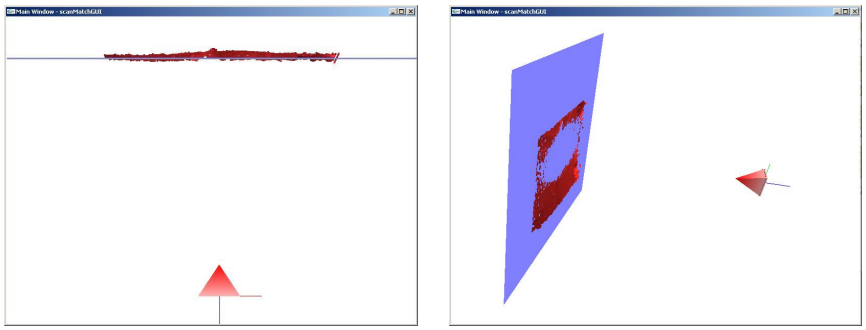

Figure 3: Measurements shown in Figure 2 after the application of per-pixel calibration.

a linear shift was all that was required to match a point in the range image with the equivalent points in the thermal and colour images. Correction for distance was only required for points closer than around 0.75m but points closer than 1m were often unusable due to overloading of the pixels, and were filtered out anyway. This combination of sensors provided a composite image with 7 channels per pixel – depth, near infrared reflectance, near infrared ambient intensity, red, green, blue and temperature.

In addition, for each image taken, the position of the pan-tilt unit was recorded, allowing the location of each pixel in 3D space relative to the robot to be determined. The two axis accelerometer was then used to determine the roll and pitch of the robot base so that the 3D data could be pre-rotated to the horizontal.

4 Map building

The process of map building may be broken down into the following components:

- Collecting range, colour and thermal images in a particular direction, along with annotations such as landmark and victim details – a "snap" such as in Figure 4
- Merging multiple "snaps" and corresponding annotations from the one location into a local map – a "scan" such as in Figure 5
- Combining multiple "scans" taken in different locations into one global map such as in Figure 9
- Postprocessing the map and including photographs and other data to generate intuitive displays and reports

Figure 6 also shows a photograph from an elevated perspective, showing how the 3D reconstruction corresponds to the actual location.

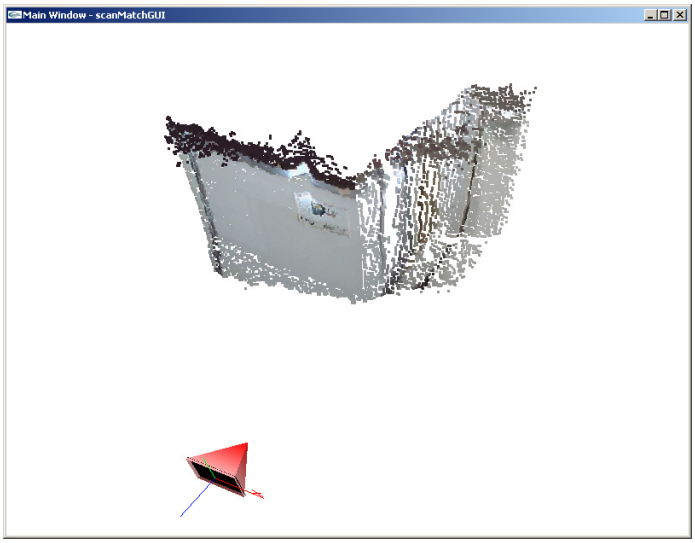

Figure 4: An example of a snap. Red arrow denotes the robot's orientation.

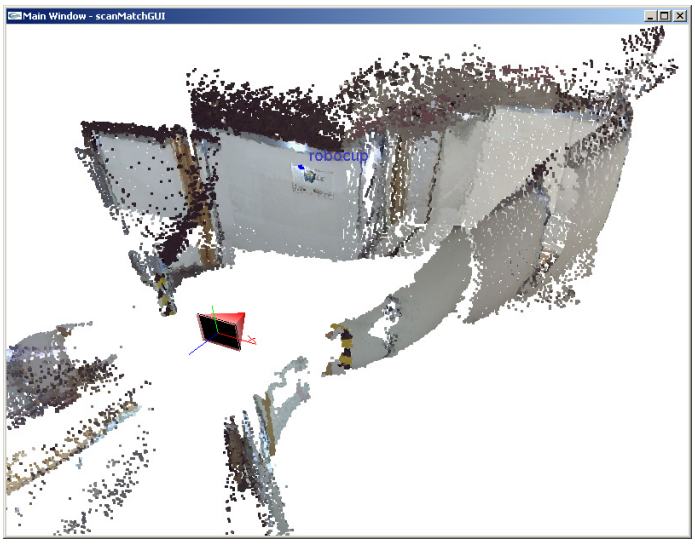

Figure 5: An example of a scan. Red arrow denotes the robot's orientation. Note the annotation of the RoboCup sign.

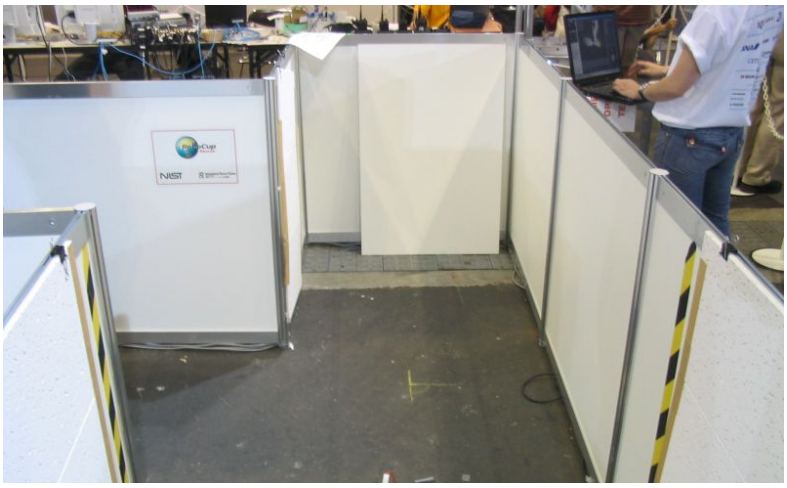

Figure 6: A photograph of the scene depicted in figure 4 and 5 from an elevated position. Note the excellent resemblance.

4.1 Snaps

A major advantage of co-locating the three sensors is that the operator can perform annotations in a highly intuitive fashion and automatically record location, visual and thermal data, which is stored in the "snap" data structure.

To indicate the presence of a landmark or victim, the operator simply clicks on the location on the display where it is observed and enters a text label into a small dialog box. The corresponding location in the range and thermal images is directly obtained, and the location in 3D space of that landmark or victim, relative to the robot, plus temperature, may be determined and recorded along with the "snap" images. It is not necessary for the operator to know or estimate how far away the victim or landmark is. This allows the annotations to follow the localised "map" generated by a single snap, the result of which can be useful even if the snap itself is mislocalised.

4.2 Scans

While the robot is stationary, the pan-tilt unit allows the operator to take many "snaps" and directly combine them into a local area map around the robot's current location, called a "scan". No correspondence matching is required as the relative directions of each "snap" is known to within fractions of a degree. Measurements from the accelerometer are then used to rotate this local area map relative to the horizon. This enables mapping to be performed even when the robot is at a steep angle.

Macro actions were developed to automate the collection of scans. These actions involve moving the pan-tilt unit around in such a way that all locations around the robot may be covered by "snaps". Generally this involves taking 10 "snaps" at intervals of 36°, stopping at each location for long enough to ensure no motion blur, a process that takes approximately 20 seconds overall. When merged using the pan-tilt unit measurements, these "snaps" form a local map

of every point in the environment that has line-of-sight to the robot.

If the operator desires, other "snaps" may also be recorded should objects of interest be above or below the ring of "snaps" taken by the macro action, so long as the robot base does not move. As these snaps are also located precisely relative to other snaps taken from the same position, landmark and victim annotations may be located relative to all the snaps in the scan, producing a very accurate and informative local area map.

4.3 Global Map

Each "scan" is pre-rotated based on the robot's roll and pitch so it is only necessary to move "scans" in 4 dimensions – X,Y,Z and yaw – to register them with the global map. Odometry is too poor on CASTER – especially given the uneven surface – for use in matching since the robot skid steers and often operates on surfaces that can shift. In fact, it was found that an odometric motion model would often be a poorer estimate of position than a zero-motion model, primarily due to large intermediate errors in heading and the high probability of not having full traction.

Variations on the Iterative Closest Point (ICP) [6] algorithm were investigated in order to register scans. This algorithm finds correspondences between points of an incoming *data set* and an existing *model set* by assuming that a given point in one set corresponds to the closest point in the other. The two sets are then transformed to minimise the least-squared error between corresponding points. The process iterates until some stopping condition, generally based on the reduction in error between iterations, is reached. Various heuristics may be applied to cull points that cannot be matched, break correspondences that can be eliminated based on distance or other factors, and weight correspondences.

Unfortunately, being a gradient descent algorithm, ICP suffers from local minima. Each pair of scans involves around 400,000 to 500,000 3D points, of which around 10% correspond to depth edges and other significant features. Thus techniques that attempt to address local minima issues, such as simulated annealing, become very expensive. Also, there must be significant overlapping regions between scans, otherwise ICP tends to match globally consistent features, which are invariably the ring of floor pixels around each scan. Thus while two scans may be taken at a reasonable distance, the local (and indeed, global) minimum might in fact be the two scans on top of each other.

As the operator takes scans, landmarks are recorded as described in Section 4.1. The range imager allows the system to locate the 3D position of these landmarks relative to the robot. If two or more landmarks in one scan match landmarks in the existing model, in theory it becomes possible to exactly register the new scan with the existing model (the extra degrees of freedom are accounted for since the robot's tilt and roll are known from the accelerometer). This system was implemented and was effective as long as the operator marked landmarks accurately.

Unfortunately, this process is time consuming from the operator's perspective as they must point the pan-tilt unit and input the landmark label. It was found that it was faster the scans to be manually registered with a user interface than it was to specify enough scans and landmarks for automated matching. Indeed, in some cases there was no overlap between scans so no automatic matching based on scan data alone could be used. Therefore, for competition

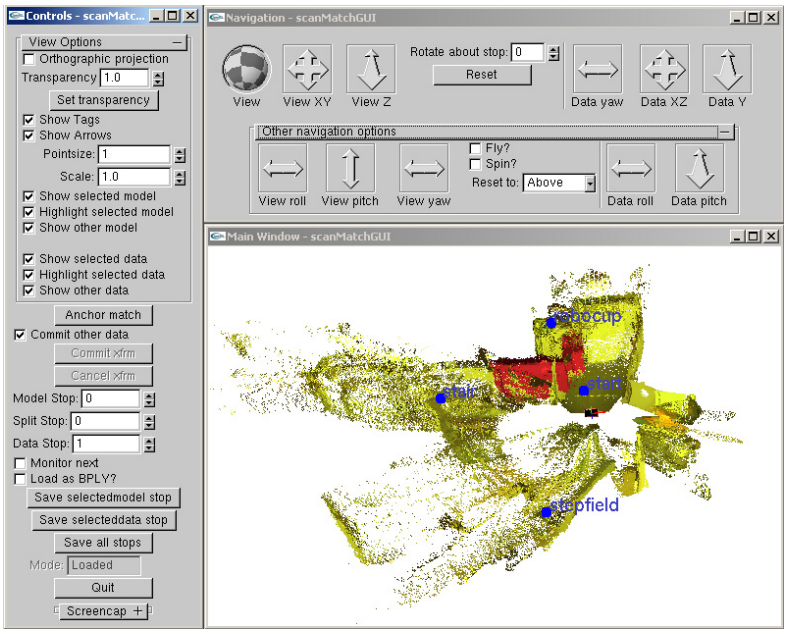

Figure 7: The user interface used for manual scan registration

runs, where the 10-minute time limit resulted in the operator only being able to take 3 or 4 scans per run, manual scan registration was preferred. The interface used for performing manual scan registration is shown in figure 7.

4.4 Synthesis and presentation to incident commander

Once the scans are registered, another software tool called RescueVis allows the operator to view the report and data on victims. It also allows the printing of a report that shows where the victims are and images of the victims. This map and victim report is then handed to the Incident Commander for scoring. The user interface for RescueVis is shown in Figure 8.

Figure 9 shows a typical map generated for the Incident Commander using the 3D capabilities of RescueVis. Subsequent screens also show pictures of the victims as well as nearby landmarks.

5 Evaluation and Further Work

The maps generated by our system were regarded as amongst the best in the competition. The ease with which the operator could place annotations with exceptional accuracy was particularly useful. Combined with the ability to directly observe the form of features in the environment in the 3D maps, this resulted in Furthermore, our score per victim, which is dependent on local map quality, was ten per cent higher than any other team. This indicates that our limitation was not the mapping, but the mobility of the robot itself; other robots beat us because they were able to find more victims, not because they produced

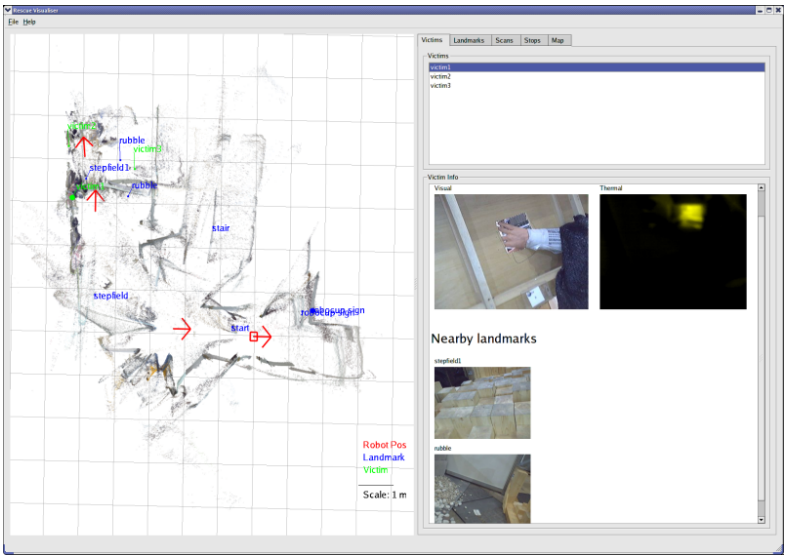

Figure 8: User interface showing victim browsing in RescueVis

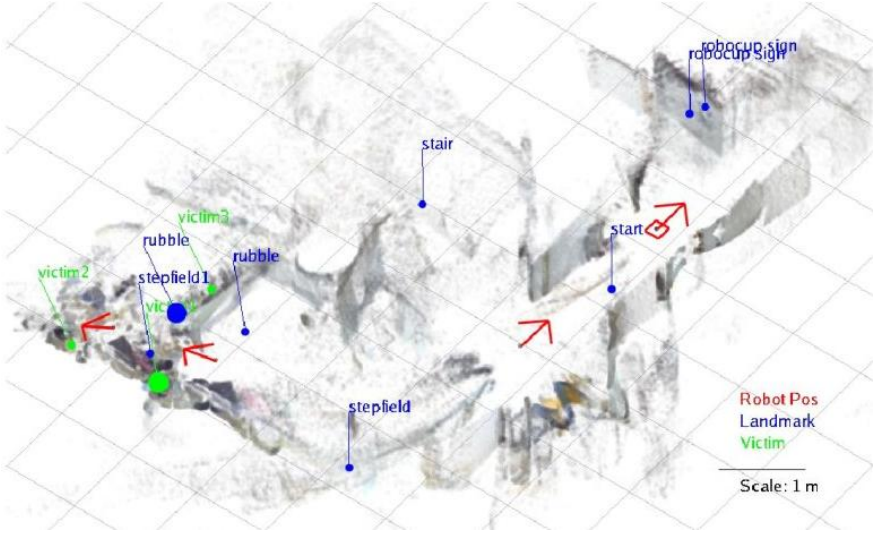

Figure 9: Generated 3D map for of Round 1 of the Preliminaries of Robocup Rescue 2005

higher quality maps.

Despite this success, there were several issues with the use of the range imager. Unfortunately a combination of insufficient strength in the pan-tilt unit, combined with communications lag, resulted in the time taken for a "scan" to at times reach 30 seconds (ideally it should take less than 5). Thus, significantly fewer "scans" could be taken than would have been desirable. Improvements in the strength of the pan-tilt unit and streamlining of the data management aboard the robot should allow this process to be speeded up.

The positioning of the range imager on the same pan-tilt unit as the driving camera was advantageous in allowing annotations to be easily made but carried the disadvantage that the operator could not observe the surroundings easily while a "scan" was being made. The addition of multiple pan-tilt cameras, as some teams have done, may alleviate this issue.

Matching adjacent "scans" is still a topic of further work. Clearly manual matching is not a desired long term solution to this problem. The ability to take more "scans" by speeding up their collection may be augmented by the addition of more sensors, such as a precision IMU or a flat laser line range scanner, such as a Hokuyo URG. Whilst only effective when the robot is horizontal, the line range scanner can operate continuously and, combined with data from the accelerometer and a magnetometer, may still be useful for tracking the robot's approximate position and providing a starting point for point based matching.

6 Conclusion

We have developed a system that effectively generates dense, textured and accurately annotated 3D maps of indoor, unstructured environments. The system was effectively deployed aboard a Robocup Rescue Robot League robot and the maps generated were pivotal in Team CASualty achieving 3rd place.

This mapping system was based on a sensor that has only recently become widely available – a range imager, in the form of the CSEM SwissRanger SR-2. By obtaining a distance measurement for every pixel, the need to scan the environment to obtain a single depth image is eliminated, although scanning is still needed if an extended field of view is required. The ability to easily merge this data with other image data, such as colour and thermal images, allows for very rich sets of data with known correspondences to be obtained, and enables automatic placement of annotations in 3D space.

Problems encounted in the use of this sensor include variations in measurements across the sensor, depth aliasing beyond 7.5m and focal blurring. These problems were solved or worked around successfully.

For a variety of reasons, automated data registration was not successful so an interface for manual scan registration was implemented. Several improvements to the mapping system are proposed that can help to solve the problem of automated scan registration for this application.

The sensor we used was one of the first compact 3D range imagers. While the sensor clearly has flaws (such as the depth aliasing, calibration and focus issues), it still presents new opportunities for constructing 3D maps in unstructured environments that would otherwise be unavailable. Its potential as a tool for both teleoperative mapping and autonomous control is very promising; and it is something we plan to explore fully.

Acknowledgment

We would like to thank firstly the Australian Research Council for its funding of the ARC Centre of Excellence for Autonomous Systems. We would also like to thank our team partners at the University of Technology Sydney for the use of their USAR test facility, in particular Jonathan Paxman and Jaime Valls-Miro. Finally we would like to thank NIST, and in particular Adam Jacoff and Brian Weiss, for organising the 2005 RoboCup Rescue competition.

References

[1] CSEM. CSEM Swiss Ranger SR-2 Applications, 2005.

[2] Hokuyo. Hokuyo URG series, 2005.

[3] Andrew Johnson. *Spin-Images: A Representation for 3-D Surface Matching*. PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, August 1997.

[4] Lingyun Liu and Ioannis Stamos. Automatic 3D to 2D Registration for the Photorealistic Rendering of Urban Scenes. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005.

[5] T. Oggier, M. Lehmann, R. Kaufmann, M. Schweizer, M. Richter, P. Metzler, G. Lang, F. Lustenberger, and N. Blanc. An all-solid-state optical range camera for 3D real-time imaging with sub-centimeter depth resolution (SwissRanger). In *Optical Design and Engineering. Edited by Mazuray, Laurent; Rogers, Philip J.; Wartmann, Rolf. Proceedings of the SPIE*, volume 5249, pages 534–545, February 2004.

[6] Hartmut Surmann, Andreas Nuchter, Kai Lingemann, and Joachim Hertzberg. 6D SLAM – Preliminary Report on Closing the Loop in Six Dimensions. In *Proceedings of the 5th IFAC Symposium on Intelligent Autonomous Vehicles*, 2004.

[7] Sebastian Thrun. Robotic Mapping: A Survey. Technical Report CMU-CS-02-111, School of Computer Science, Carnegie Mellon University, 2002.

[8] Yujin. Robhaz dt3 web site, 2005.