Improving Vision-Based Maps By Using Sonar and Infrared Data *

MARK A. LANTHIER
School of Computer Science,
Carleton University
Ottawa, Ontario, Canada
email: lanthier@scs.carleton.ca

DORON NUSSBAUM
School of Computer Science,
Carleton University
Ottawa, Ontario, Canada
email: nussbaum@scs.carleton.ca

AN SHENG
IMAGO Machine Vision Inc.
5 Laval St, Suite 203
Hull, Quebec, Canada
email: asheng@imagotrackers.com

ABSTRACT

Vision-based sensors such as stereo cameras, are often used on mobile robots for mapping and navigation purposes. Cameras provide a rich set of data making them useful for object recognition, localization and detecting environmental structure. When obtaining range measurements, however, stereo camera vision systems do not perform well under some environmental conditions such as regions which are uniform in appearance (e.g., plain walls), large metallic or glass surfaces (e.g., windows) and poor lighting conditions. This paper describes how range data obtained from a stereo camera vision system can be improved upon through use of additional sonar and infrared proximity sensors. We provide experimental results showing that data fusion from three types of sensor range data does indeed result in a more accurate occupancy grid mapping.

KEY WORDS

Stereo Camera, Sonar, Infrared, Data Fusion, Mapping

1 Introduction

The commonly known issues of sensor noise and sensor inaccuracies have prompted much research in the area of sensor fusion [1][2][3][4][5]. The abilities and limitations of various sensors often dictates their usefulness for certain robotic applications. Sonars, for example, are often confined to collision avoidance tasks rather than mapping because of the low spatial resolution and various noise problems which are due to crosstalk and specular reflection. Many researchers have turned toward 3D laser range finders and passive stereo camera vision systems to obtain a more accurate representation of the environment. Laser range finders are still fairly expensive and so many robots are equipped with less accurate stereo camera vision systems. Such vision systems do provide fairly accurate range measurements but there are still some situations in which they cannot provide adequate distance measurements such as windows and very black mat materials [6]. Some of the ranging problems encountered with stereo cameras can be overcome through making assumptions on typical environmental structure. However, in our research, we are interested to show how multiple heterogeneous sensors can improve the acquired vision range data without any assumptions on the environmental structure.

Sensor fusion has been a subject for much research [1] using various methods such as Kalman filtering [2], Bayesian reasoning [3], artificial networks [4] and fuzzy logic [5]. When doing indoor mapping and navigation, researchers have found that the use of different types of sensors on a robot can be beneficial due to their varying characteristics. Work has been done to fuse sonar sensor data both with infrared sensor data [4] and with laser range finder data [6][7][8], as well as fusing camera data with sonar data [9][10]. Although camera or laser vision systems are typically used to obtain fine environmental details, cheaper sensors such as sonars and infrared sensors are often used to provide a quick and rough estimate of the environment [7]. These cheaper sensors can also be used to compliment and confirm the readings of the vision systems. Wilhelm et. al. [9], for example, show how sonar data provides an additional weight which is used to increase the probability of tracking a human when using camera vision data.

The work presented here represents our initial test results aimed at identifying where stereo camera vision systems fail when mapping indoor environments and showing that data from other inexpensive sensors (in our tests, ultrasonic transducers (i.e., sonars) and infrared proximity sensors) can be fused with the vision data to improve the accuracy. Although the sonars and IR sensors are not without their own limitations, we show that they can be used to accurately fill in missing/invalid data from a typical stereo camera system as well as help to identify noisy vision data. Also, while most research in data fusion concentrates on just two types of sensor data, in this work we concentrate on merging data from three types of sensors. In this preliminary work, our experiments focussed on building maps in a static environment. Moreover, we do not address issues of localization or navigation, but instead focus on heterogeneous sensor fusion from fixed known positions in the environment.

2 Robot and Sensors

Figure 1 shows the robot used in our experiments with a closeup view of the stereo camera, infrared sensor array

*Research partially supported by NSERC.

and sonar ring. Note that since our sonars and infrared sensors are in fixed positions, our experiments concentrated on performing data fusion on data obtained from a particular fixed height in the environment. A similar approach was taken by [6].

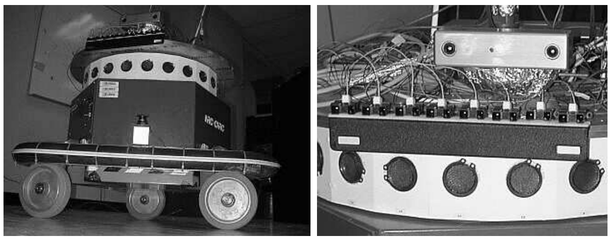


Figure 1. The robot used for testing and a closeup view showing the BumbleBee stereo camera, infrared proximity sensors and sonar ring.

Stereo Cameras

Our robot is equipped with a passive two-lens stereo vision camera system [1] as can be seen in Figure 1. Through software, the cameras are able to extract 3D distance information from a single robot position at a rate of 10Hz. The cameras can accurately determine distances within a range of 2m to 10m with a field of view at around 45°, allowing a large area of the environment to be processed from a single image. One disadvantage of stereo cameras is that they require adequate lighting and even differences between natural and artificial lighting can pose a problem when extracting fine details [9]. This is not a problem in typical indoor environments during normal operating hours but can become an issue for nighttime applications with reduced lighting conditions. A more serious drawback of this sensor is its inability to detect environmental features when there is an absence of detail in the scene image or when obstacles have transparent or metallic surfaces. For example, plain (i.e., uniform colour) walls provide no detail in which to make matches between pixels in the stereo images.

Although this type of sensor produces accurate distance data for typical indoor environments, it often leaves "gaps" in which no distance measurement is available. Additionally, in the case of windows or large metallic surfaces, invalid range data is often returned which could lead to invalid mapping of obstacles. A typical approach to tackling these issues is to examine the full 3D range of data obtained from the stereo camera and make assumptions on the environmental structure. We aim however, to reduce these gaps and discrepancies from the stereo camera range data at a specific height through use of sonars and IR sensors.

Sonar Ring

Our robot also hosts a ring of 24 Polaroid 6500 ultrasonic transducers (i.e., sonar sensors) as can be seen in Figure 1. It is widely known that sonars have relatively high distance accuracy and range (10m) but low angle resolution because of the wide beamwidth (i.e., $\pm 15^\circ$). The wide beam is unable to distinguish features within the beam angle, making sonars a poor choice of sensor for fine feature extraction within indoor environments. This resolution problem is magnified for objects further away from the robot (i.e., objects appearing at the wide end of the beam). Unfortunately as well, the well known problems of specular reflection, unwanted echoes, and crosstalk can cause invalid or imprecise range readings. Often, by taking multiple readings from various locations, the sonars can produce a fairly accurate, yet coarse, representation of the environment. Sonars provide the greatest benefit when used as a compliment to other sensors to reconfirm the presence of obstacles. Furthermore, since they are not vision-based, they are useful under poor lighting conditions or when there are many transparent objects such as windows or glass doorways, as this is where traditional vision-based sensors fail.

Infrared Proximity Sensor Array

Lastly, our robot is also equipped with an array of 8 Sharp GP2Y0A02YK infrared proximity sensors as can be seen above the sonar ring and below the cameras in Figure 1. A main drawback of these sensors is that they are close-range sensors and can only accurately measure obstacle distances within a range of 0.1m to 1.5m. Thus many readings must be taken in order to determine the shape of larger obstacles. Another drawback of these sensors is that they are susceptible to inaccuracies due to outdoor light interference as well as an obstacle's colour or reflectivity characteristics which can be seriously affected by windows and metallic surfaces. In terms of precision however, IR sensors can be more accurate than the sonars and the stereo camera when detecting larger shaped features such as walls, doorways, desks, cabinets etc.

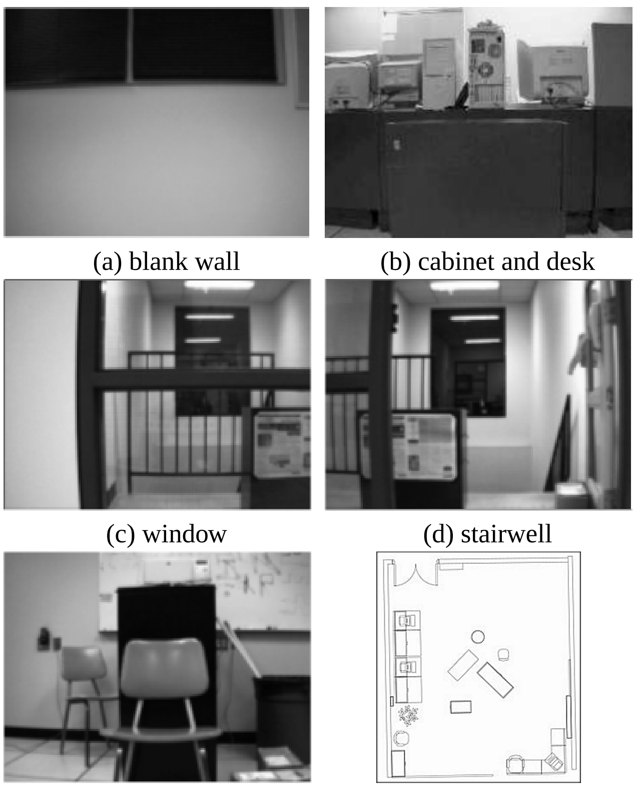
3 Data Fusion

As mentioned previously, the aim of our experiments was to show how sonar and IR detector data readings can be combined with the stereo camera data to improve the overall indoor map at a specific fixed height in the environment. A series of experiments (described in the subsections to follow) were designed to address three important issues. First, since the stereo cameras perform poorly on uniformly coloured regions, there is a need to show how the combining of sonar and IR data can "fill in" the missing gaps in the camera data. Second, it is beneficial to show how the data fusion of IR and sonar data can help eliminate noisy range data from the camera data, thus helping to refine the environment's contours. Lastly, since vision-based sensors fail under transparency and low-lighting conditions, we need to show how the sonars

[1] *Bumblebee* from www.ptgrey.com/products/bumblebee

can help under such conditions.

Figure 2 shows images depicting our test cases. For our window/stairwell test, Figure 2.c and 2.d show the left and right halves of the scene where a hallway window next to an open doorway peers into a stairwell. Figure 2.e shows the full lighting scene for our low lighting condition test (as the low lighting image would be difficult to see in print). Finally, Figure 2.f shows a diagram of a large portion of our lab used in our full lab test. For each of our tests, we show the scaled range data overlayed on top of a diagram of the environment. The diagrams show the shape of the environment as well as the robot's positions from which the range data was taken (shown as a circle with a rectangle indicating direction of camera and IR sensors).



(a) blank wall (b) cabinet and desk

(c) window (d) stairwell

(e) low lighting setup (f) diagram of lab

Figure 2. Images depicting the test cases for our experiments.

There are various sensor fusion strategies for producing occupancy grids. We perform data fusion on our occupancy grid using a one dimensional linear Kalman filter [2]. Figure 3 shows how we determine which occupancy grid cells are updated from a single sonar reading. The obstacle in the figure is detected within the sonar's emitting angle of $30°$ with an error rate of 1% (shown exaggerated in the figure for clarity). The diagram shows the probability distribution across the grid cells which must be updated as a result of this reading. This distribution is a combination of the sensor dependant angular and distance distributions that are used during the data fusion process. We use a Gaussian distribution to compute the angular distribution in modelling the sensor noise. Therefore the innermost grid cells within the emitting angle have a higher probability of occupancy than the outermost cells. We also apply this strategy to compute the distance distribution in modelling the sensor distance noise. The combined distribution is shown in the figure as a greyscale occupancy grid where the darker shaded cells indicated a high probability of object occupancy and lighter shaded cells represent a low probability of occupancy. When using IR and stereo camera sensors, we apply the same probabilistic model but the angular distribution is not applied since it becomes insignificant due to the much smaller beam emitting angle.

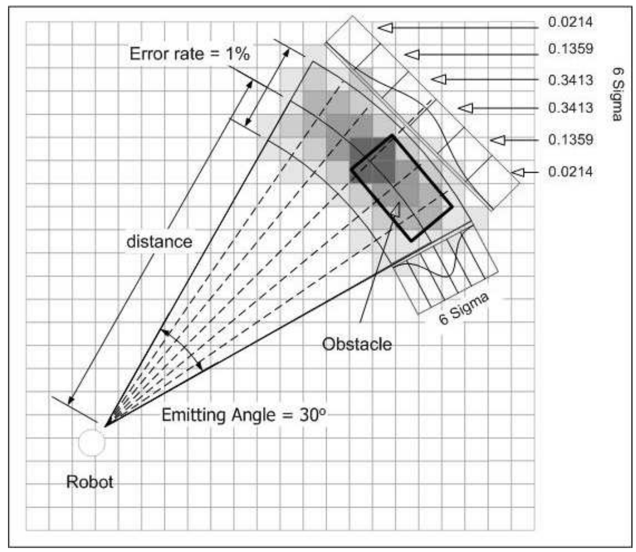
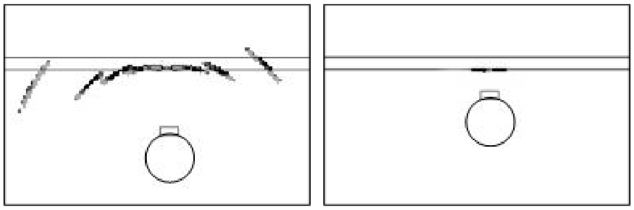


Figure 3. Diagram showing the occupancy grid cells that need to be updated during a single sonar reading.

3.1 Uniformly Coloured Regions

Our first test of the vision system was on a blank wall (see bottom half of Figure 2.a). With this uniformly coloured wall, the stereo cameras were unable to produce any valid range data whatsoever. The sonars and IR detectors however, were able to obtain fairly accurate range readings for the wall as can be seen in the diagrams of Figure 4. Note that the sonar data was obtained from a single position, hence showing the angular nature of the sonar measurements. This however can be made more linear through the fusing of multiple readings from different positions. When combined for mapping purposes, these two sensors indicate a solid obstacle presence and can be fused to fill in the absent range data from the stereo cameras. This is more evident in our full lab test of section 3.4.

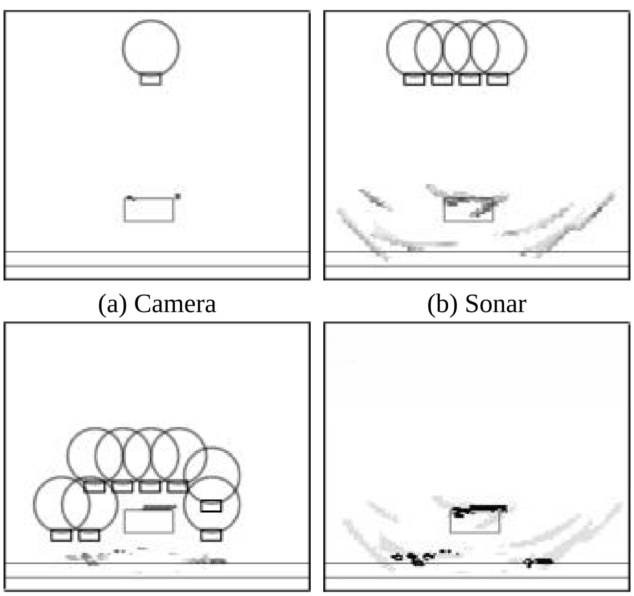
Our next test was on the scene in (Figure 2.b). In the bottom half of the image, there is a small grey rolling cabinet that blends in with the grey back panel of the desk



(a) Sonar (b) IR

Figure 4. Range data from the blank wall test.

containing the computers. The contrast between the cabinet and the back desk panel is very low, where only the sides and top of the cabinet are distinguishable. The diagrams of Figure 5 show the range results. Notice that the stereo camera was able to accurately detect the left and right edges of the cabinet, indicating that the camera had no problems with the low contrast between the cabinet and desk. As in the blank wall test, the range data for the surface of the cabinet and desk panel was unobtainable using the camera. Both the sonar and IR sensors were able to detect the differences between the cabinet and desk panel. The image in Figure 5.d shows a classification of the sensor readings where the black cells indicate agreement in readings between two or more sensors, and the light grey cells indicate spurious measurements which cannot be trusted. Note that the sonar and IR data sets were both necessary in order to obtain an accurate distance to the cabinet and desk.

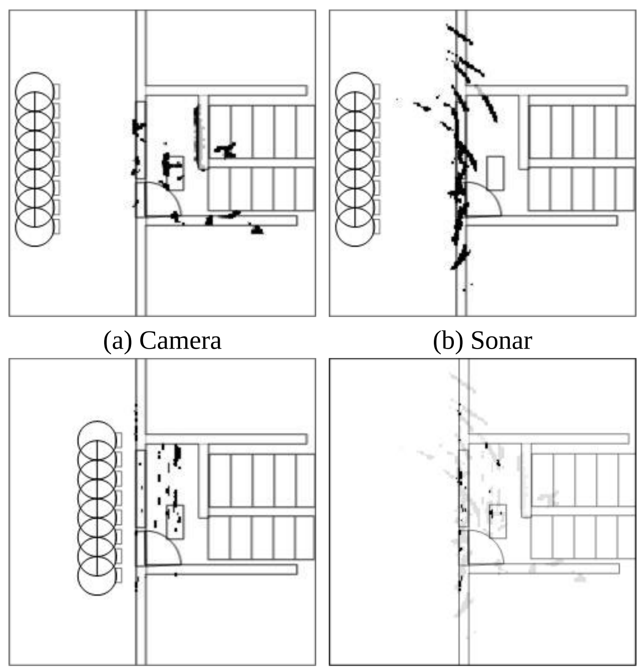


(a) Camera (b) Sonar (c) IR (d) Classification

Figure 5. Range data from the cabinet/desk test.

3.2 Transparent Regions

Another test performed was that of the scene in (Figure 2.c and 2d). These are two images that when pieced together form the scene. The left image depicts a large window which overlooks a stairwell and the right image shows an opened door to that stairwell. A large cabinet was placed in the stairwell to provide additional features to the test scenario. The diagrams of Figure 6 show the range results. Notice that the data from the stereo camera shows detected ranges to the stairwell railings and cabinet through the window. It was able to correctly detect the door frame and the wall through the opened door, but is unable to properly detect the window distance. The sonar readings show that the window is detected although the angular variation inherent to the sonar's ring arrangement is clearly evident. Notice that the IR data is also affected by the transparent window, where the railing is being detected as well as the cabinet. Although the IR sensor detected these features, the range data is inaccurate; likely since some of the IR light is being reflected back while some is passing through. In the classification image (d), we can see that very few cells show agreement between two or more sensors, although most cells in agreement are along the window and due to the availability of sonar data. It is clear from this test that without the use of sonar, the window would be virtually undetectable.

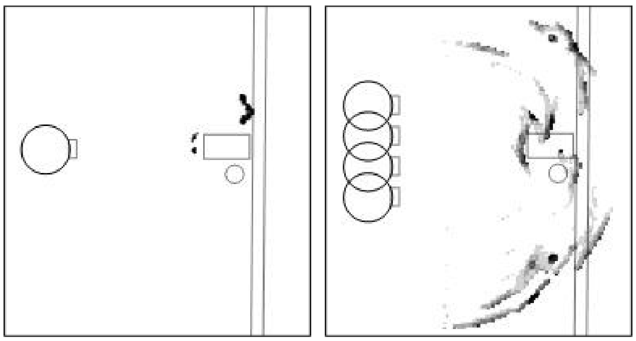


(a) Camera (b) Sonar (c) IR (d) Classification

Figure 6. Range data from the window/stairwell test.

3.3 Low Lighting Conditions

In order to verify that stereo cameras have poor performance in a low lighting setting, another test was performed with obstacles arranged as shown in Figure 2.e. The images in Figure 7 show the range results for this test. Notice that the stereo camera had difficulties detecting the dark garbage can, cabinet and walls while the higher contrast chairs were still partially detectable under the low lighting conditions. The range readings, however, do not provide adequate information for detecting obstacles, making the vision sensor of very little use in low lighting scenarios. From just a few position readings, the sonar is able to detect important features, although quite coarse. The sonar has difficulty with the rounded garbage can. Nevertheless, the important walls are detected. These readings may also be verified and fine-tuned through the addition of IR data. From this test it is clear that the sonars provide an advantage over the stereo cameras under low lighting conditions.



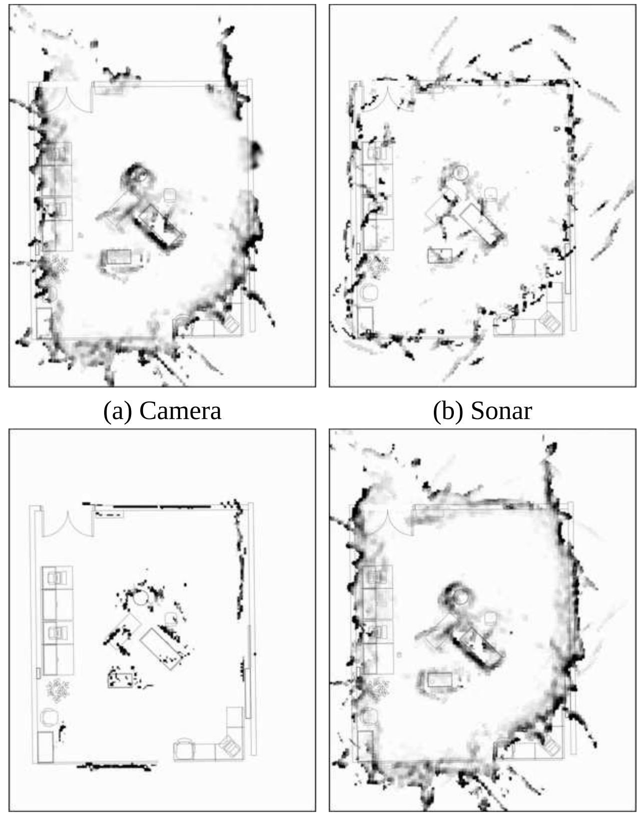
(a) Camera (b) Sonar

Figure 7. Range data from the low lighting test.

3.4 Full Lab Map

To get a better sense of the practicality of our 3-sensor data fusion, we performed an experiment which produced a map (occupancy grid) of a large portion of our lab. The results are shown in Figure 8. In our lab, we do not have any windows at the height with which the range data was taken. Also, we performed the test under normal lighting conditions. This test therefore shows the way in which the data fusion can fill in any range gaps due to regions of uniform colour and also how the overall map can be fine tuned. For each type of sensor, the robot was moved to various locations in order to obtain a full coverage of the test environment. Notice in the stereo camera data that the top and right regions show gaps. These are caused by the large regions of uniform colour on the walls in the lab. The top left gap represents an open doorway and the "splatter" effect is caused by range information to objects in the corridor outside the lab. The bottom of the grid has a splatter effect as well since there are additional obstacles beyond the measured portion of our lab.

The sonar data provided a reasonable outline of the environment and performed better than the camera in the uniform regions. It did however produce more spurious readings and had some difficulty with the finer details for the objects in the center of the lab. The IR data shows a very precise reading of the walls, with some missing data on the left and bottom right because of desks which prevented the robot from getting close enough to the walls



(a) Camera (b) Sonar

(c) IR (d) Combined

Figure 8. Range data from the full lab test.

to take readings. The occupancy grid obtained through the data fusion as shown in the combined image, clearly indicates the advantage of the sonar and IR data sets. Notice how much more precise the map is along the top and right borders where the IR data helped to refine the missing stereo camera data. Also, notice how the three sensors combined provide a more accurate shape estimate for the inner obstacles.

A more thorough analysis of the sensor data will help us to see the benefits of having more than two sensors. One simple measure of showing the benefit is to examine the occupancy grid cells for which each sensor produced range readings. Figure 9.a shows an image which classifies grid cells for the full lab test according to the number of sensors that gave readings for that cell. The light grey cells represent the 59.7% of the total readings in which only one sensor produced a range value. There were 34.2% of the readings confirmed by two sensors (shown in medium grey). Only 6.2% of the readings were confirmed by all three sensors (shown in black). This classification indicates which environmental features are commonly detected by sensors with different characteristics and shows that multiple heterogenous sensors can produce complimentary data. Figure 9.b depicts the map represented by the combined data fusion of Figure 8.d where all single sensor classification data readings have been removed. Hence about 60.4% of the data is discarded as untrustworthy. Now examining the data from the stereo camera, Figure 9.c shows (as circled) where the sonar and IR range data were able to fill in the gaps that were unavailable from the vision data due to regions of uniform colour. These regions represented a significant 12% of the lab's border, which would have otherwise been void of range data if only the stereo camera data was used. Lastly, Figure 9.d shows the "noise" (a significant 48.8%) that was eliminated from the stereo camera data through the data fusion process.

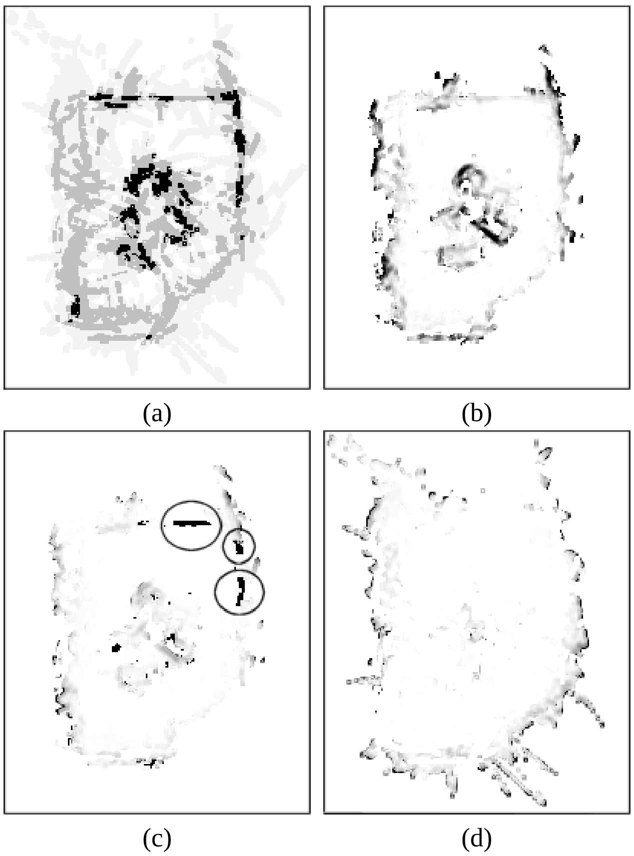


Figure 9. (a) Grid cell weight classification, (b) Combined data with single sensor readings removed, (c) Vision data combined with Sonar/IR data showing where gaps were filled in, (d) Noise that was eliminated from the vision data.

4 Conclusion

In this research we investigated the need for multiple sensors on a robot for mapping unknown environments. We focussed on the ability of the sensors to detect the range to objects which are uniform in appearance, objects of different materials (such as transparent objects) and objects under different conditions (e.g., low lighting). Through our experiments presented here, we have clearly shown the limited abilities of stereo camera vision systems in certain scenarios. The experiments indicate that the use of low cost sonar and infrared proximity sensors can be used in combination to accurately "fill in" the missing vision sensor information (up to 12% of the border in our full lab test) and also to fine-tune existing vision sensor range data by discarding noise (48.8% in our full lab test). We plan to further our work by applying probabilistic methods to the fusion of the individual sensor's occupancy grids so as to reduce the false readings. We also plan to investigate how sensors can be dynamically given priority over others so that the "best" sensor is used when operating under various conditions.

References

[1] R.R. Brooks and S.S. Iyengar, *Multi-Sensor Fusion* (Prentice Hall, New Jersey, USA, 1998).

[2] G. Welch and G. Bishop, *An Introduction to Kalman Filter*, University of North Carolina, Chapel Hill, TR 95-041, May 23, 2003.

[3] P. Sykacek and I. Rezek, *Markov Chain Monte Carlo Methods for Bayesian Sensor Fusion* (Kluwer Acadamic Publishers, 2000).

[4] Humberto Martínez Barberá, Antonio Gómez Skarmeta, Miguel Zamora Izquierdo and Juan Botía Blaya, Neural Networks for Sonar and Infrared Sensors Fusion, *3rd Intl. Fusion Conference (FUSION 2000)*, Paris, France, 2000.

[5] M. Lopez, F.J. Rodriguez and J.C. Corredra, Fuzzy Reasonning for Multisensor Management, *1995 IEEE Int. Conf. on Systems, Man and Cybernetics*, Vol. 2, New York, USA, 1995, pp. 1398-1403.

[6] Ashraf Aboshosha and Andreas Zell, Robust Mapping and Path Planning for Indoor Robots based on Sensor Integration of Sonar and a 2D Laser Range Finder, *Proc. of the IEEE 7th Int. Conf. on Intelligent Engineering Systems*, Assiut-Luxor, Egypt, 2003.

[7] Gregory Dudek, and Paul Freedman and Ioannis M. Rekleitis, Environment Exploration Using "Just-in-Time" Sensor Fusion, *Vision Interface*, Toronto, 1996, pp. 175-182.

[8] Moshe Kam, Xiaoxun Zhu and Paul Kalata, Sensor Fusion for Mobile Robot Navigation, *Proceedings of the IEEE*, 85(1), 1997, pp. 108-119.

[9] T. Wilhelm, H.-J. Böhme and H.M. Gross, Sensor Fusion for Vision and Sonar Based People Tracking on a Mobile Service Robot, *Proc. Int. Workshop on Dynamic Perception*, Bochum, 2002, pp. 315-320.

[10] B.R. Duffy, C. Garcia, C.F.B. Rooney and G.M.P. O'Hare, Sensor Fusion for Social Robotics, *31st International Symposium on Robotics (ISR2000)*, May 14-17, 2000, Palais des congrs, Montral, Canada.