Detection of Small Obstacles at Long Range Using Multibaseline Stereo

Todd Williamson and Charles Thorpe

Abstract— **This paper presents a multibaseline stereo technique specially suited to detecting obstacles. First, we describe traditional multibaseline stereo and explain why it is not suitable for obstacle detection. This analysis leads directly to a suitable method. Then we describe the obstacle detection system that we have built. Finally, we show results from our system for a variety of different obstacles under different lighting conditions.**

Keywords— **Obstacle detection, Trinocular stereo, Automated Highway Systems, Multibaseline stereo.**

I. Introduction

Highways present dynamic environments with real-time constraints. A vehicle travelling at highway speeds needs to detect obstacles at ranges of 100 meters in order to swerve or brake. Sensors such as automotive radar do not have the acuity to find small obstacles at this distance, and have significant difficulties with non-metallic obstacles such as wood, cement, or animals. While a variety of competing methods have been proposed for on-road obstacle detection, most of the work has focused on detecting large objects, especially other vehicles (e.g., [3]). Many of these methods can successfully detect moving vehicles, but the more difficult problem of finding small, static road debris such as tires or crates remains unsolved. Although the problem of detecting static obstacles has been tackled in both the cross-country and indoor mobile robot navigation literature (e.g., [4]) these systems have operated at low speeds (5-10 mph) and short range.

In this paper, we present an obstacle detection system that is capable of detecting small static obstacles on the order of 15cm tall on the road surface up to 100m ahead of the vehicle. First we discuss traditional stereo techniques, and why they do not work well in a highway environment. Then we present our solution to the problem, followed by a description of the system we have implemented. The results of using our system to detect a variety of obstacles under a variety of different conditions are then shown.

II. Traditional Stereo

As illustrated in Figure 1, traditional stereo processing involves taking two images of a scene at the same time from different viewpoints. Each point in one of the images is constrained by the camera geometry to lie along a line (called the *epipolar* line) in the other image. Its position on this line is related to the distance of the point from the cameras.

In order to make the search more reliable, instead of comparing individual pixels from the two images, small regions are compared.

Two examples of this are shown in Figure 1. In these images, the scene is of the inside of a garage. The garage door has calibration targets attached to it to enhance image texture, and the images have been filtered to enhance image texture. Two regions are chosen as examples of stereo matching.

For the example region on the door of the garage, we see that the regions searched in the stereo matching (shown in detail below the images) match very well. The upper curve on the graph at the bottom shows the matching error (sum of absolute differences, SAD) as a function of the displacement along the epipolar line. This graph shows a strong global minimum at the correct value of 100.

On the other hand, the example on the garage floor does not match as well. This is due to the fact that, since the ground is tilted with respect to the camera axis, points which are higher in the image are actually farther away and thus match at a different location. This is seen as a difference in the slope of the line on the ground. The lower curve of the graph shows that the global minimum of the matching error does not occur at the correct position (which would be at a value of around 155).

It is clear from this example that a simple application of traditional stereo techniques will not be sufficient for detecting obstacles on a road surface; points on the ground such as those shown in the example will produce incorrect results, particularly in regions where the image texture is low. Since the problem is caused by a difference in the geometry of the surfaces being observed, the solution to this problem is to compensate for the different geometry.

III. "Ground Plane Stereo"

The simplest way to solve the problems described in the previous section is to warp one of the images (using a

T. Williamson and C. Thorpe are with the Robotics Institute, Carnegie Mellon University, Pittsburgh PA 15213. Email: {Todd.Williamson, Charles.Thorpe} @ri.cmu.edu.

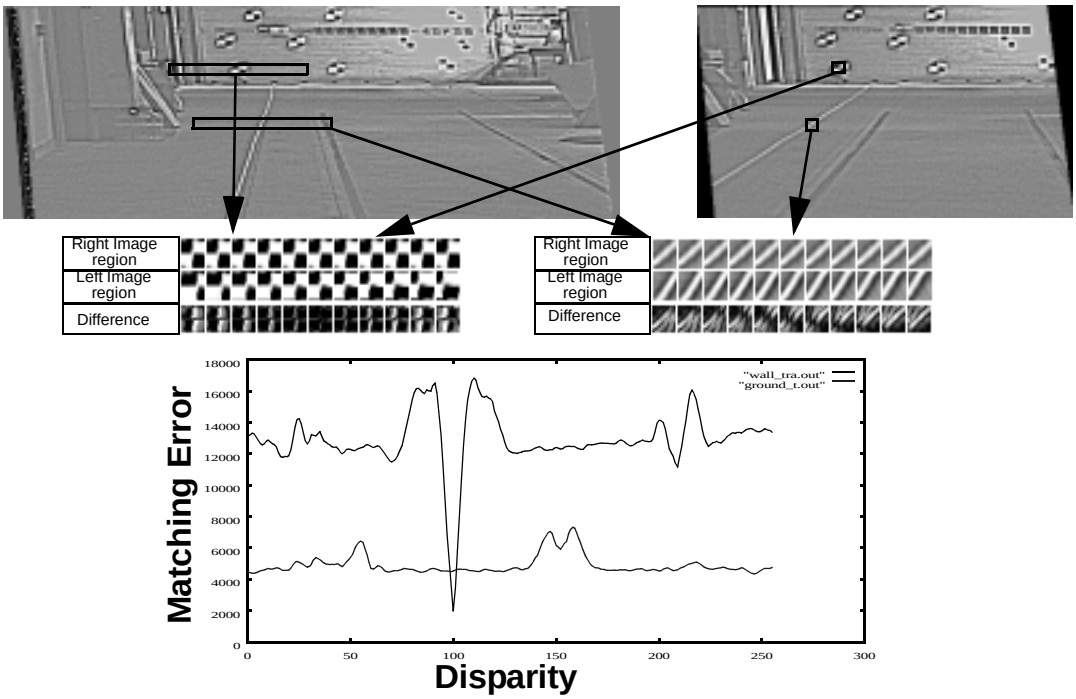


Fig. 1. Traditional Stereo Processing

projective warping function) so that the images would appear to be exactly the same if all of the pixels in the image were on some typical ground plane. This results in a situation as shown in Figure 2. Both images now appear to be the same for pixels which are on the ground, but pixels which are on a vertical surface such as the wall of the garage are now warped in much the same way that the ground pixels were warped in traditional stereo. This means of computing stereo (described in more detail in [6]) is effectively the same as the "tilted horopter" method of Burt et al.[1].

Comparing the results from the ground plane method with the results from the traditional method, we notice several differences. First, the global minimum of the matching error curve for the point on the ground (the lower curve) now appears at the correct location. The value of the error at the minimum is also lower than before, since it matches better. Second, although the global minimum of the curve for the point on the door is still at the correct location, the trough of the minimum is much wider, indicating a less certain result. The value at the minimum is larger, indicating that it does not match as well.

This example illustrates an interesting result: if we compute stereo using both methods, it is possible to determine whether a given point lies on a vertical surface (if the traditional method produces at lower minimum error) or on a horizontal surface (if the ground plane method produces a lower minimum error). The correct disparity can also be determined from the position of the lower minimum.

Since most obstacles that we are concerned with contain nearly vertical surfaces, detecting such obstacles becomes very easy using this method.

One issue that must be addressed is what conditions are necessary for this method to work reliably. For example, if two surfaces appear in the same image region (near where the garage door meets the ground, for instance), which surface will be chosen? This most important factor is the magnitude of the image texture on each surface. Another factor is how close the surface directions are to being vertical or horizontal.

Figure 3 shows the results of applying both methods to a typical input image set. The gray coding in both cases represents the number of pixels of displacement along the epipolar line (dark is negative, medium gray is zero, and bright is positive). As expected, the ground plane method does very well on the ground pixels, but poorly on the wall in the background. Conversely, the traditional method works well on vertical features such as the lamp post and the wall, but there is a lot of noise on the ground surface.

IV. Implementation

Figure 4 shows the architecture of the system that we have implemented. Three CCD cameras with 35mm lenses are arranged in a triangular configuration, mounted on top of our Toyota Avalon test vehicle. The distance between the outer set of cameras is about 1.5m.

The computation that is performed is based on that used by the CMU Video Rate Multibaseline Stereo

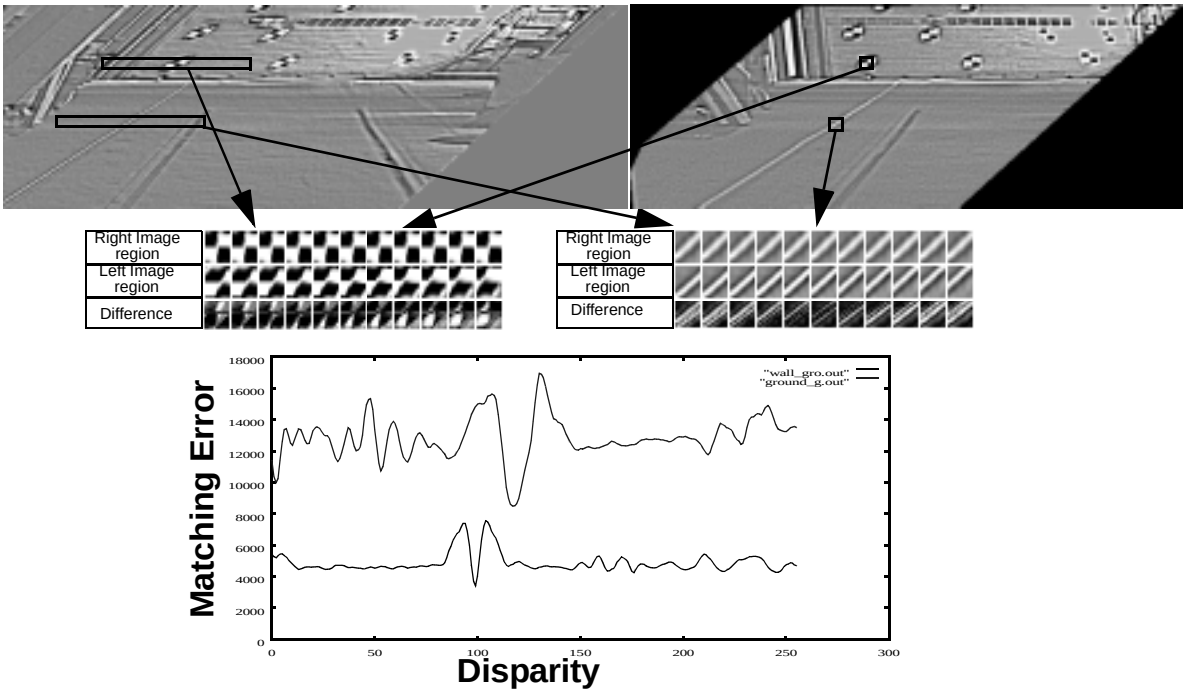
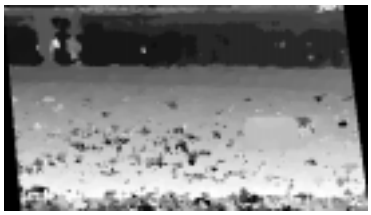
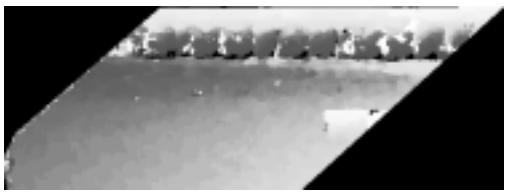


Fig. 2. “Ground Plane Stereo”

Original Image



Traditional Method Output



Ground Plane Method Output

Fig. 3. Example output of both methods

Machine[2]. The images are first passed through a Laplacian of Gaussian (LoG) filter, then rectified to align the epipolar lines. Stereo matching is then performed using both the traditional method and the ground plane method. Based on the output of both methods, the further step of obstacle detection and localization is performed.

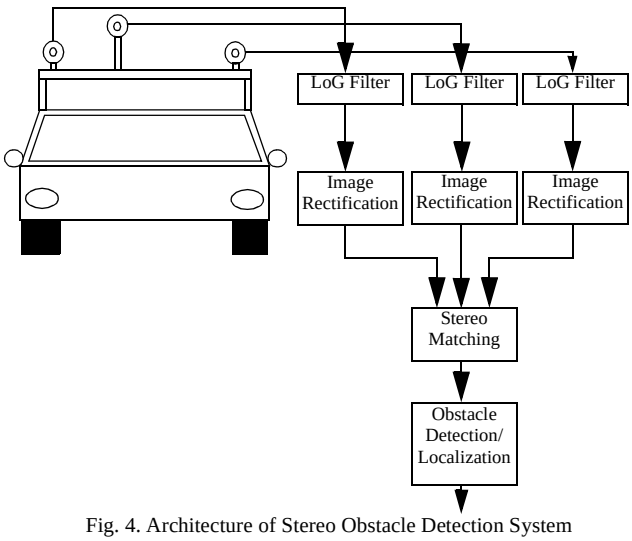


Fig. 4. Architecture of Stereo Obstacle Detection System

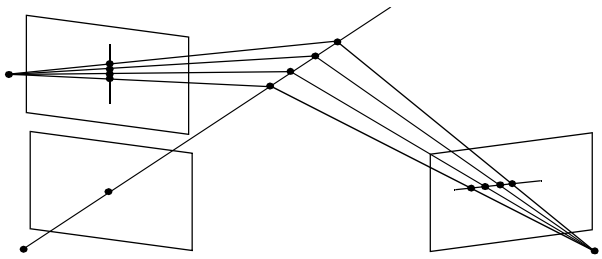


Fig. 5. Three cameras in an “L” configuration give different epipolar directions

A. Multibaseline (Trinocular) Stereo

There are several benefits to adding a third camera in a triangular configuration. The most important of these is that the epipolar lines for different pairs of cameras are in different directions (as illustrated in Figure 5). This is due to the fact that the epipolar direction is the same as

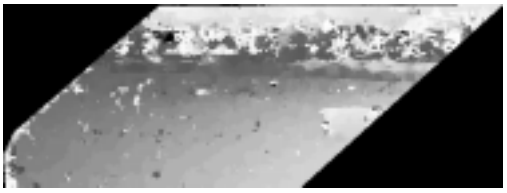


Fig. 6. Example using only two cameras

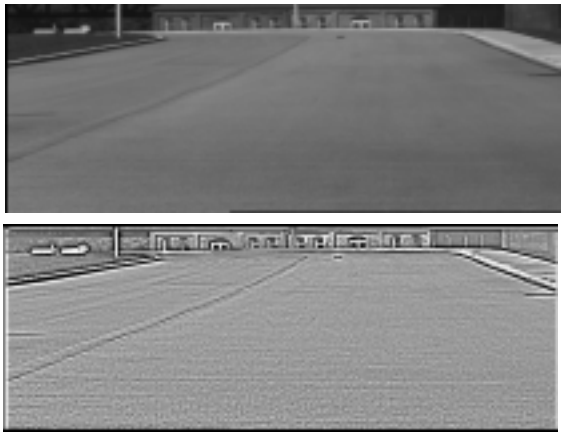


Fig. 7. Image before and after LoG filtering

the direction of displacement between the cameras. This is important in situations where the image has texture in one direction but not in the other (for example, the top border of the obstacle in Figure 3).

Another benefit of adding additional cameras is that it allows multiple measurements at each point. This is useful in increasing accuracy and rejecting noise. Furthermore, a system containing only two cameras can be confused by repeated patterns in the image (such as lines painted on the road surface). With three cameras, this problem is eliminated.

Adding a fourth camera does provide some additional benefit, but it becomes much more difficult to perform the stereo matching efficiently.

Figure 6 shows the output for the ground plane method from Figure 3 if only two cameras are used. The number of incorrectly matched pixels is much larger than with three cameras.

B. Laplacian of Gaussian (LoG) Filtering

Laplacian of Gaussian filtering is a well-accepted means of extracting features to match from multiple cameras, while at the same time compensating for differences in camera gain and bias. We use an LoG filter with a high gain in order to enhance the texture of the otherwise featureless gray asphalt. The results of this filtering are shown in Figure 7. The increase in image texture is very apparent.

The importance of the LoG filter to our algorithm is illustrated in Figure 8. The lack of image texture on the road surface causes the entire region to be unmatchable,

Fig. 8. Example of stereo output without LoG filter

Fig. 9. Detected vertical surfaces

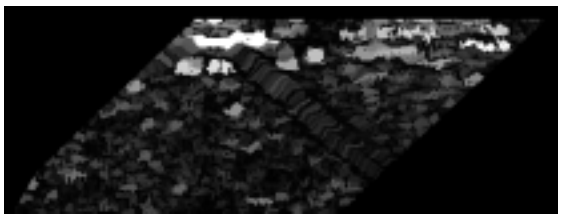


Fig. 10. Size of regions of constant disparity

although regions with higher texture, such as the obstacle itself and the curb, are still computed correctly.

C. Obstacle Detection from Stereo Output

As discussed in Section III, our method involves performing two types of stereo matching (for vertical and horizontal surfaces), and comparing the absolute errors to determine whether a particular image region belongs to a vertical or horizontal surface. The result of this is shown in Figure 9. The regions shown in the lower image are coded by the size of the difference between the minimum errors. Brighter regions indicate that the vertical match is much better than the ground plane match. Thus, regions which appear white are most likely to be vertical, and black regions are most likely to be horizontal.

Regions of very low texture (such as the white stripe down the side of the road) sometimes match well as vertical surfaces because of differences between the individual cameras being used.

In order to remove such false obstacles from consideration, we use a very simple confidence measure. For regions which are actual vertical surfaces, we expect that the traditional stereo matching method will return a relatively large number of pixels at approximately the

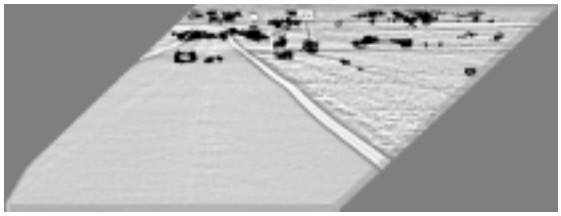


Fig. 11. Detected Obstacles

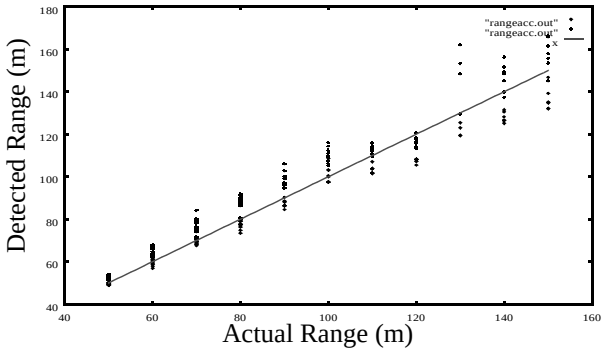


Fig. 12. Stereo range accuracy

same depth. Conversely, if a region belongs to a horizontal plane, we would expect the traditional method to report a number of different depths. Using standard connected components labeling methods on the disparity image generated from traditional stereo matching, we get the image shown in Figure 10. This image encodes the size (in pixels) of the region to which each pixel belongs. Large regions appear brighter, and these regions are more likely to be obstacles. By requiring detected obstacle regions to pass this consistency check, we can remove most false positive detections.

Combining the images of Figure 9 and Figure 10, we get the detected obstacle output shown in Figure 11. Obstacles are shown in black. This example shows two 14cm high obstacles, which are pieces of wood painted white and black. The obstacles are 100m in front of the vehicle.

V. Results

We have collected a set of test data using wooden obstacles of four different heights (9, 14, 19, and 29cm) and three different colors (white, black, and gray) at measured distances from 50 meters to 150 meters.

Figure 12 shows the accuracy of the detected range for all 12 obstacles. As expected, the measured range is very accurate when the object is close, and becomes increasingly less accurate as the obstacle gets farther away.

The results of running the obstacle detection system are shown in Table I. This table shows that we were successfully able to detect obstacles that are bigger than 9cm at up to 110m.

Table I:
Obstacle Detection Results

	Black 9cm	Black 14cm	Black 19cm	Black 30cm	Grey 9cm	Grey 14cm	Grey 19cm	Grey 30cm	White 9cm	White 14cm	White 19cm	White 30cm
50m	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
60m	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	✓
70m	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
80m	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
90m	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	✓
100m	✓	✓	✓	✓	×	✓	✓	✓	✓	✓	✓	✓
110m	✓	✓	✓	✓	×	✓	✓	✓	×	✓	✓	✓
120m	✓	×	✓	✓	✓	×	✓	✓	×	✓	✓	✓
130m	×	×	✓	×	×	✓	✓	✓	×	×	×	×
140m	✓	×	✓	✓	✓	✓	✓	✓	×	×	×	×
150m	✓	✓	✓	✓	✓	✓	✓	✓	×	×	×	×

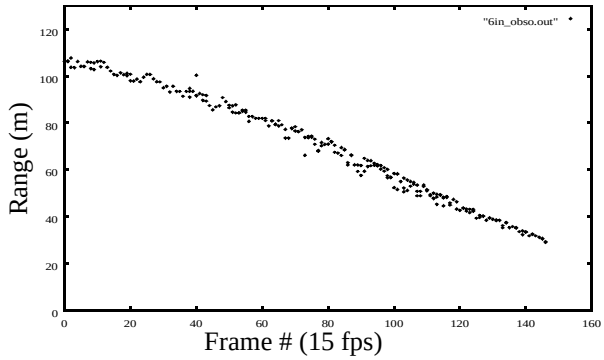


Fig. 13. Detection trace for 14cm obstacle

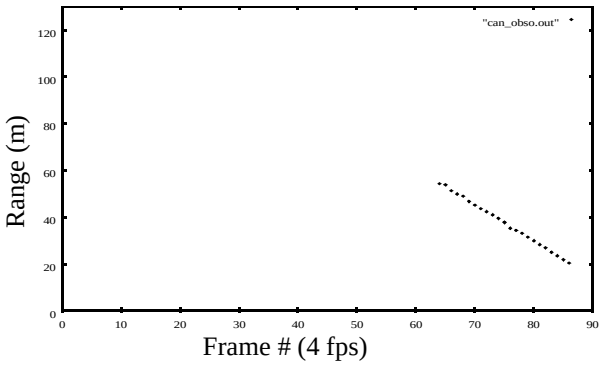


Fig. 14. Detection trace for soda can

Figure 13 shows an example trace of an obstacle detection run. The vehicle moved at a constant rate (about 25 km/h) toward a 14cm black obstacle of the type shown in Figure 9. The data was taken at 15 fps, and processed off-line. The obstacle is detected in every frame of the data, out to a maximum range of approximately 110m (which is the beginning of the data set).

Figure 14 shows the same type of trace, this time for a standard 12oz (350ml) white soda can. The soda can is first reliably detected at 57m.

Each of the previous examples has shown only the detections that actually represented the obstacle. Of course, there are many more detected objects. A full trace is shown in Figure 15, along with an example image from the set and a diagram showing an overhead

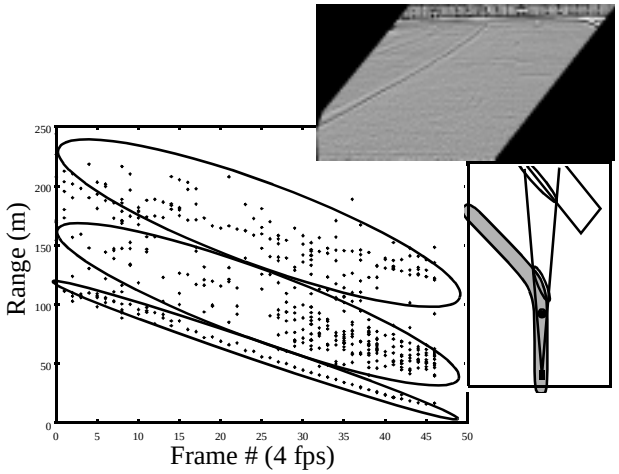


Fig. 15. Other detected points

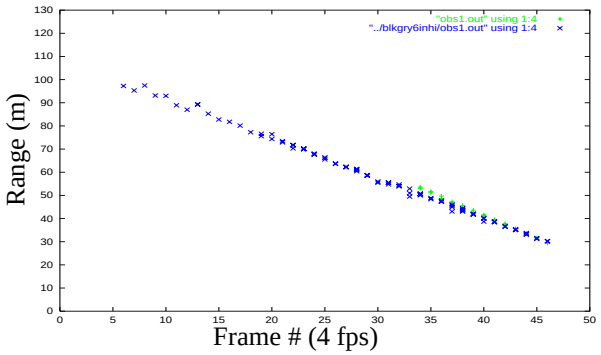


Fig. 16. Detection trace for a 14cm black obstacle at night. x represents data taken with high beams, + are with low beams.

view of the scene. The detections can be divided into three sets, representing the obstacle, the curb behind the obstacle, and the building in the background. Also note that there are no false detections that are closer than the obstacle.

TABLE II:
RESULTS OF REPEATED EXPERIMENTS

	<30 m	30-40m	40-50m	50-60m	60-70m	70-80m	80-90m	90-100m	100-110m	110-120m	120-130m	>130m
total frames	86	50	71	70	85	81	93	92	119	91	48	62
frames detected	86	50	71	70	84	81	85	76	99	60	28	32
percent detected	100	100	100	100	98.8	100	91.3	82.6	83.2	65.9	58.3	51.6

Table II shows the results of ten repeated experiments in detecting a 9cm black obstacle. The images from each run were logged to disk, and then hand classified into bins based on the range to the obstacle. The number of frames falling into each bin is listed in the second row of the table. Of this set of frames, the number which were detected properly is shown in the third row of the table. The percent detection is shown in the last row. As can be seen from the data, the system is able to detect even a 9cm tall obstacle reliably out to a distance of 80m, after which the reliability degrades gracefully.

Figure 16 and Figure 17 show results for detecting obstacles at night. Figure 16 shows results looking at a white obstacle, both with high beam headlights and low beams. The obstacle is detectable at nearly 100m with high beams, but that is reduced to 60m with low beams. Figure 17 shows the results for a black obstacle. As is to be expected, the black obstacle is harder to detect, at 60m for high beams and 40m for low beams.

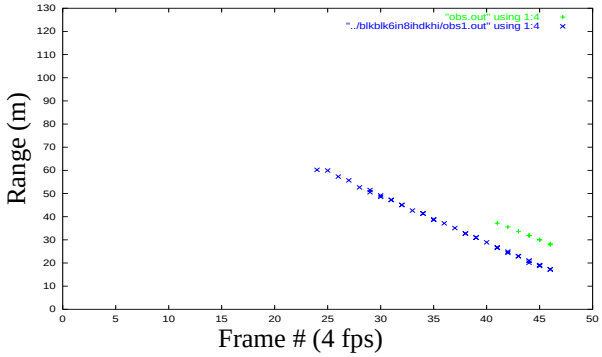


Fig. 17. Detection trace for a 14cm black obstacle at night. x represents data taken with high beams, + are with low beams.

VI. CONCLUSIONS

We have demonstrated an obstacle detection system that uses trinocular stereo to detect very small obstacles at long range on highways. The system makes use of the apparent shape of surfaces in the image in order to determine whether pixels belong to vertical or horizontal surfaces. A simple confidence measure is applied to reject false positives introduced by image noise. The system is capable of detecting objects as small as 14cm high at ranges well in excess of 100m.

VII. ACKNOWLEDGMENT

This research was sponsored by the collaborative agreement between Toyota Motor Corporation and Carnegie Mellon University.

VIII. REFERENCES

[1] P. Burt, L. Wixson, and G. Salgian. "Electronically Directed "Focal" Stereo," *Fifth International Conference on Computer Vision (ICCV '95)*, Cambridge, MA, June 1995, pp. 94-101.

[2] T. Kanade, A. Yoshida, K. Oda, H. Kano, and M. Tanaka. "A Stereo Machine for Video Rate Dense Depth Mapping and its New Applications," *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR'96)*, June, 1996.

[3] Q.T. Luong, J. Weber, D. Koller and J. Malik, "An integrated stereo-based approach to automatic vehicle guidance," *Fifth International Conference on Computer Vision (ICCV '95)*, Cambridge, Mass, June 1995, pp. 52-57.

[4] L. Matthies and P. Grandjean. "Stereo Vision for Planetary Rovers: Stochastic Modeling to Near Real-Time Implementation," *International Journal of Computer Vision*, 8:1, pp. 71-91, 1992.

[5] L. Robert, M. Buffa, M. Hébert. "Weakly-Calibrated Stereo Perception for Rover Navigation," *Proceedings of the International Conference on Computer Vision (ICCV '95)*, 1995.

[6] T. Williamson and C. Thorpe. "A Specialized Multibaseline Stereo Technique for Obstacle Detection," *Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR '98)*, June, 1998.