Stochastic Performance Modeling and Evaluation of Obstacle Detectability with Imaging Range Sensors [1]

Larry Matthies Pierrick Grandjean
Jet Propulsion Laboratory - California Institute of Technology
4800-Oak Grove Drive
Pasadena, California 91109

Abstract

Statistical modeling and evaluation of the performance of obstacle detection systems for Unmanned Ground Vehicles (UGV's) is essential for the design, evaluation, and comparison of sensor systems. In this paper, wc address this issue for imaging range sensors by dividing the evaluation problem into two levels: quality of the range data itself and quality of the obstacle detection algorithms applied to the range data. Wc review existing models of the quality of range data from stereo vision and AM-CW LADAR, then used these to derive a ncw model for the quality of a simple obstacle detection algorithm. This model predicts the probability of detecting obstacles and the probability of false alarms, as a function of the size and distance of the obstacle, the resolution of the sensor, and the level of noise in the range data. Wc evaluate these models experimentally using range data from stereo image pairs of a gravel road with known obstacles at several distances. The results show that the approach is a promising tool for predicting and evaluating the performance of obstacle detection with imaging range sensors.

1 Introduction

In semi-autonomous, cross-country navigation of UGV's, obstacle detection is achieved primarily by processing data from imaging range sensors. For example, both scanning laser range-finders (LADAR) and stereo vision systems have been used in demonstrations of cross-country navigation by prototype UGV's. To date, sensor limitations have forced these demonstrations to usc relatively large obstacles (e.g. $\geq$ 50 cm) and low driving speeds (e.g. $\leq$15 km/h). To complement research on improving the sensors, there is now a need for quantitative methodologies for modeling and evaluating sensor performance. Such methodologies will bc useful for measuring progress, for system design, and for comparing competing technologies. Although such methodologies have been developed *in* other domains, they have not yet been applied to obstacle detection sensors for UGV navigation.

To bc more specific, wc address obstacle detection from range imagery alone, where range images arc 2-D arrays of 3-D range measurements produced by LADAR, stereo vision, or other closely related sensors. For this problem, performance modeling and evaluation has questions at two levels: (1) how good is the range data itself and (2) how good are the obstacle detection algorithms that are applied to the range data? Ideally, wc would like to answer these questions with a common methodology that applies to LADAR, stereo vision with daylight or thermal cameras, and possibly other sensors.

The quality of range data can be evaluated in terms of random and systematic errors, Theoretical models of random errors have been developed for both LADAR [1] and stereo vision [2, 3, 4, *51*. Systematic errors, or biases, have been examined experimentally for LADAR in [6] and for stereo in [7]. For the quality of obstacle detection, a relevant theoretical framework exists for modeling the probability of detection and false alarms in classical detection problcms [8]. This framework has been applied to obstacle detection with LADAR for a spacecraft landing application [9]; however, no work has been done yet on applying this to obstacle detection for ground vehicles.

In this paper, wc focus on obstacle detection with range images produced by stereo vision. Section 2 elaborates on our motivation and overall methodology for designing and evaluating obstacle detection sensor systems. Section 3 addresses the quality of range data by reviewing stochastic models for random errors in both stereo and LADAR range imagery, then by conducting an experimental evaluation of random and systematic errors in range images produced from stereo images of gravel roads with known obstacles at various distances. Section 4 conducts an analogous development for the quality of obstacle detection, For a simple obstacle detection algorithm wc have used in real-time, outdoor demonstrations, this section derives models for the probability of detection and false alarms, then evaluates the models experimentally using the range imagery produced in section 3. On the design side, the results show promise for being able to usc probabilistic performance models to predict error rates for different sensor parameter choices. On the experimental evaluation side, the results show that our methodology is

[1] This work was supported by rhc Defense Advanced Research Projects Agency (DARPA) and the National Aeronautics and Space Administration (NASA).

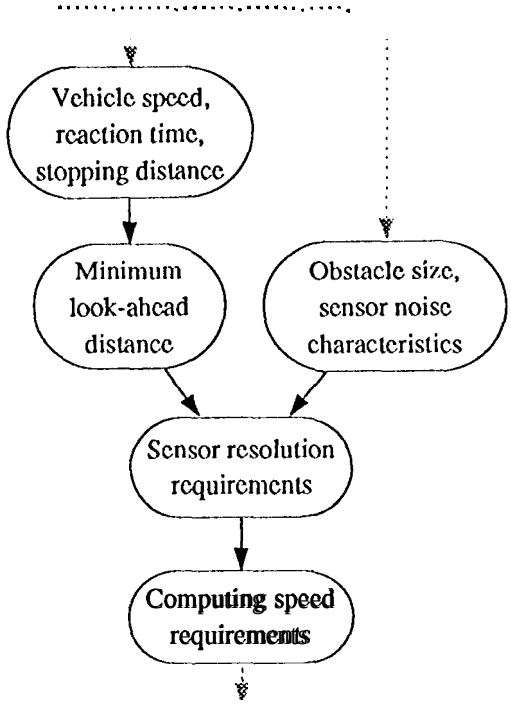


Figure 1: Overall methodology

a useful 1001 for quantitatively measuring the performance of obstacle detection with imaging range sensors.

2 Motivation and Methodology

To motivate our approach, wc will briefly consider the larger context of using range sensors for obstacle detection for UGV's. Given that wc want to drive at a given speed and detect obstacles of a given size, lhc basic questions arc (1) how far ahead do we have to detect those obstacles and (2) what sensor resolution do wc need to detect obstacles at that distance? The first question boils down to reaction time; it can be answered by estimating the distance required to stop the vehicle in case of an obstacle, given the initial velocity, deceleration rate, and latency limes for perception and actuation. In [10], wc have taken this approach to show that a minimal bok-ahead distance is:

$$d_l = d_c + v_0(2t_c + t_a) + v_0^2/(2a) \quad (1)$$

where v_0 is the initial velocity, a *is* the deceleration rate, t_c and t_a arc latency times for perception and actuation, respectively, and d_c *is* the distance the cameras arc mounted behind the nose of the vehicle.

The second question has elements of both geometry and statistics, because wc must take into account both the angular resolution of a pixel and the uncertainty introduced by noise. In purely geometric terms, it is easy to derive how many pixels are subtended by an obstacle; however, noise makes obstacle detection a stochastic process, and geometry alone does not capture the stochastic , element. Thcrcforc, reliability must be defined in probabilistic terms, for example by designing the vision system to achieve a specified detection probability.

Our overall methodology is illustrated in figure 1. In this paper, wc consider the right branch of the figure by deriving theoretical models of obstacle detectability and by evaluating the validity of these models experimentally. As stated earlier, the end goal of this work is to estimate the sensor resolution required to achieve given levels of reliability in obstacle detection. As shown in figure 1, the resolution requirement translates into computational rcsourcc requirements needed to process the imagery within given latency limits. Wc have addressed the left branch of the figure, dealing with look-ahead distance and computation requirements, in [1 O]. There wc have also analyzed how all of these factors arc related and shown how to control perception to optimize the velocity/reliability trade-off,

In the balance of this paper, wc first consider how sensor noise affects the quality of the range data, then address how noise in the range data affects the quality of obstacle detection.

3 Quality of Range Data

Wc address this issue specifically for stereo vision; analogous results for LADAR arc given in [6]. We will begin by briefly reviewing our stereo algorithm, then discuss mathematical models of performance and experimental evaluation of the validity of those models.

3.1 Stereo Matching Algorithm

The algorithm is described in detail in [1 1]; here wc review the characteristics relevant to later sections on performance modeling and evaluation, The algorithm has the following stages:

1. input stereo image pairs arc transformed into bandpass image pyramids by a difference-of-Gaussian type pyramid transformation.
2. Image similarity is measured by computing the sum-squared-diffcrcncc (SSD) for 7 x 7 windows over a fixed disparity search range.
3. Disparity is estimated by finding the SSD minimum independently for each pixel.
4. Confidence measures are computed by estimating posterior probabilities for the disparity estimates at each pixel; where the probability falls below a preset threshold, the pixel is marked as having no disparity estimate.
5. Sub-pixel disparity estimates are obtained by fitting parabolas to the three SSD values surrounding the

SSD minimum and taking the disparity estimate to be the minimum of the parabola.

6. The resulting disparity map is smoothed with a 3 x 3 low-pass filter to reduce noise and artifacts from the sub-pixel estimation process.

For textured, outdoor images, this algorithm produces disparity estimates for almost every pixel in the image. The algorithm has been implemented in a real-time system that produces range images from the 60 x 64 level of the image pyramid in approximatel y onc second pcr frame. The system has been installed on two robot vehicles and used to demonstrate obstacle detection, on gravel roads and in sand y off-road terrain, at speeds up to three kilometers per hour. For reference, these systems have baselines of 25 to 35 cm and fields of view of 30 to 40 degrees.

3.2 Statistical Performance Models

For stereo vision, performance depends on many factors, including the contrast and spatial frequency of texture in the intensity signal, the noise level in the images, various artifacts introduced in the stereo matching process, and the true range to the objects in the scene. Wc need to characterize both the random errors caused by noise and the systematic errors, or biases, resulting from artifacts. Wc will develop mathematical models of the random errors, and examine both the random and the systematic errors experimentally. Wc began to examine these issues in [7]; this paper reviews and extends the previous work, then uses these results in section 4 in characterizing the quality of obstacle detection.

Distributions of random errors need to bc examined both for the estimated disparity map and for the range estimates resulting from the nonlinear operation of triangulation. In both cases, there will bc uncertainty at each pixel and there may bc correlation in the errors for neighboring pixels (i.e. the noise may not bc white),

First consider the distribution of disparity estimates at each pixel. For good estimates of solid surfaces, wc expect the distribution to bc unimodal and compact about the mean. At sub-pixel resolution, the precision (variance) of a disparity estimate reflects statistical fluctuations around the Sub-pixel mean. The precision will vary inversely with the slope of the image intensity derivative $I_x(x, y) = \partial I(x, y)/\partial x$ along the scanline [1 1]. Assuming a flat, fronto-parallel surface, an approximate model for the variance of a disparity estimate is

$$\sigma_d^2 \approx \frac{2\sigma_n^2}{\sum_{x,y \in W} [I_x(x, y)]^2}, \qquad (2)$$

where σ_n^2 is the variance of noise in the image and the denominator is a sum over the window of squared, central-difference derivative estimates. A more elaborate variance model that takes into account slanted disparity surfaces is derived in [13].

Several factors induce statistical correlations between errors for nearby pairs of pixels. These factors include: (1) the spatial filtering employed in creating the image pyramids, which correlates the noise in the low-resolution images, (2) the fact that matching windows for adjacent pixels overlap, which will induce correlation in the disparity estimates even if the image noise is not correlated, and (3) the application of the noise-rcxluction filter to the estimated disparity fields. Such correlation will affect the uncertainly of terrain slope estimates computed from local range differences. The multiple contributing factors make the correlation somewhat complicated to model theoretically; instead, wc will fit an empirical model to the experimental results prcscntcd below.

Even if disparity errors arc Gaussian, the nonlinear triangulation operation produces non-Gaussian errors in range estimates. It is customary to ignore this by modeling range as approximately Gaussian and to estimate the variance of range errors by lincarized error propagation. Wc examined this approximation in [7] by deriving the "true" distribution of range and comparing it to the Gaussian approximation. The agreement was good, except that skew in the true distribution bccomcs more pronounced as the mean disparity gets c1osc to zero. Wc conchrdc that, for present purposes, range estimates arc well modelled as Gaussian distributed, with standard deviation

$$\sigma_Z \propto \sigma_d Z^2 . \qquad (3)$$

where Z is the 3-D coordinate expressing the distance from the cameras. Note that range estimates for nearby pairs of pixels will be correlated, because of the correlation dcscribcd earlier in the disparity estimates.

For comparison, the standard deviation of range measurements made by AM-CW LADAR has been modelled as [1, 6]:

$$\sigma_Z \propto \frac{Z}{\sqrt{\rho \cos \alpha}}, \qquad (4)$$

where ρ is the surface reflectance and α is the angle of incidcncc of the laser beam on the surface. For UGV applications, α gets closer to 90 degrees the further ahead the sensor looks; therefore, this compounds the effect of the increasing range term in the numerator. Since the graphs of (3) and (4) may cross over, comparing the performance of stereo and LADAR will require calibrating the constants of proportionality for each sensor.

3.3 Experimental Evaluation

To experimentally evaluate the quality of both range estimation and obstacle detection, wc collected an ensemble of stereo pairs of a flat, gravel road with and without obstacles of known geometry. The obstacles consisted of

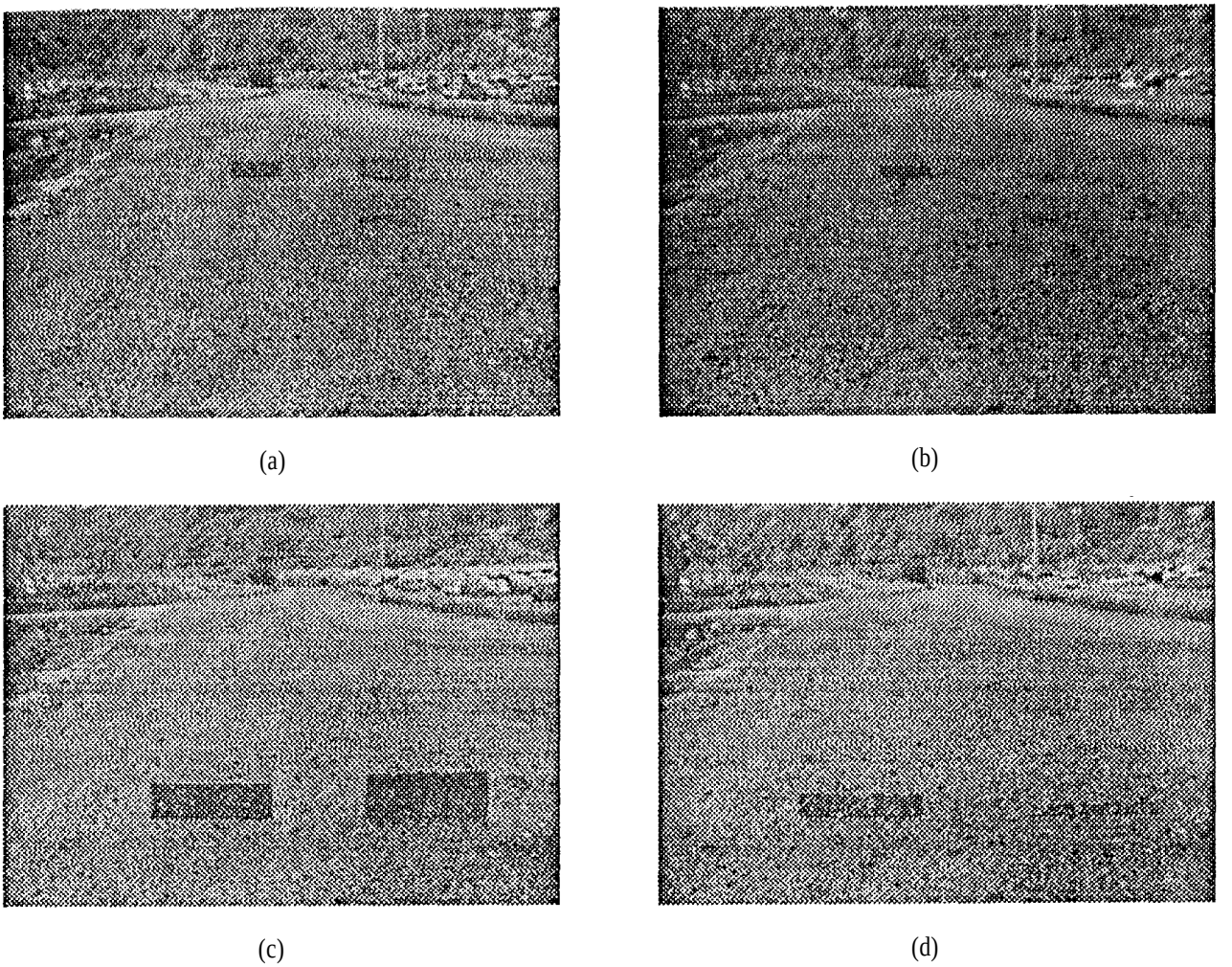


Figure 2: Test images: (a) large obstacles al 50 feet (b) small obstacles at 50 feet (c) large obstacles at 20 feet (d) small obstacle.s at 20 feet,

four rectangular boards 1 meter (m) wide by 10, 20, 30, and 40 centimeters (cm) high, which were coated with gravel from the road to provide realistic texture. Sets of 100 stereo pairs were taken with roughly constant illumination for each of the following cases: the road without obstacles, the road with the two largest obstacles (30 and 40 cm) placed at 20, 35, 50, 6S, or 80 feet, and the road with the two smallest obstacles (10 and 20 cm) placed at the same set of distances. Example images from these data sets arc shown in Figure 2. These images were processed at 60 x 64 and 120 x 128 resolution to generate sample statistics for the performance measures of interest.

3.3.1 Bias

To look for possible biases in the range estimates, wc computed the mean of the estimated disparity images for each set of 100 stereo pairs and plotted range profiles for selected columns of the mean range images. Figure 3 shows profiles for individual columns just inside the left edge of the 30 cm obstacle, for each of the five distances from the camera and for two image resolutions. Several bias artifacts arc evident. First, the total perceived height of the obstacle was always greater than the true height, and became more so the further away the obstacle was placed. Wc attribute this to a bias introduced by the finite size of the SSD window. That is, for area-based matching algorithms, disparity estimates tend to reflect the highest contrast texture in the matching window; this hinds to cause a "halo" around the obstacles within which the range estimates arc nearly the same as the range to the obstacle. Since the size of this halo is fixed in the image plane (one half the width of the matching window), its size in 3-D grows with the distance to the obstacle.. Therefore, the obstacles tend to look larger in the range image the farther

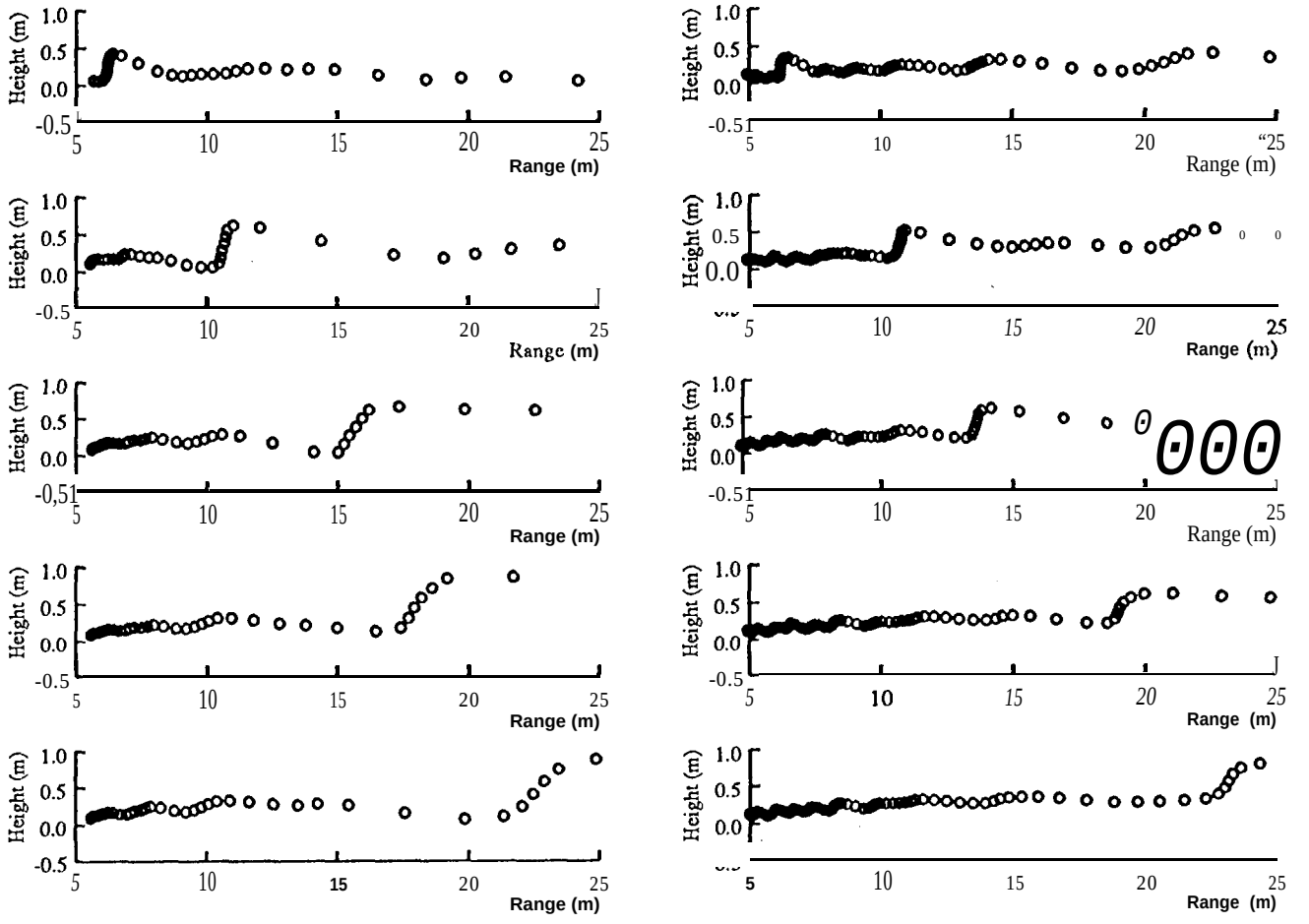


Figure 3: Range profiles for 30cm obstacle at 20, 35, 50, 6S, and 80 feet, respectively, for (a) 60x64 resolution, (b) *120x 128* resolution,

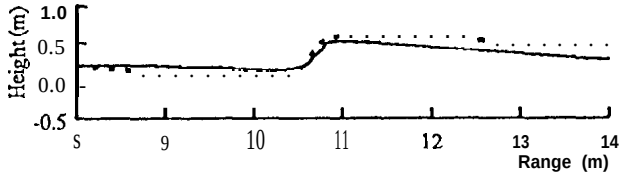


Figure 4: 60x64 (dolled) and 120x128 (solid) profiles overlaid for obstacle at 35 feet.

away they are. Since the halo is an image-plane effect, the 3-D error is reduced for higher image resolutions, as can be seen by comparing the 60 x 64 and the 120 x 128 results in Figures 3a and 3b; Figure 4 shows this more explicitly by overlaying profiles for both resolution from the 35-foot data.

Several other effects offset this halo effect, First, the obstacles appear more slanted with increasing distance, whereas in reality they were nearly vertical in all cases). This is certainly duc to their smaller size in the image. Also, at large distances the halo effect was sometimes overwhelmed by loss of contrast and texture, which have the reverse effect of making the obstacle blend in with the background.

Another type of artifact appears to be duc to biases in sub-pixel disparity estimation. The sub-pixel disparity estimation algorithm appears to be biased toward integer disparity estimates. For flat ground, this imposes a ripple pattern on the range data, which can be seen at C1OSC range in Figure 3b. In the past, wc have seen other artifacts around high contrast intensity edges [15] and around occluding boundaries of objects [7]; wc have also noticed that the amount of bias depends on the boundary texture and contrast. However, wc will not explore these here.

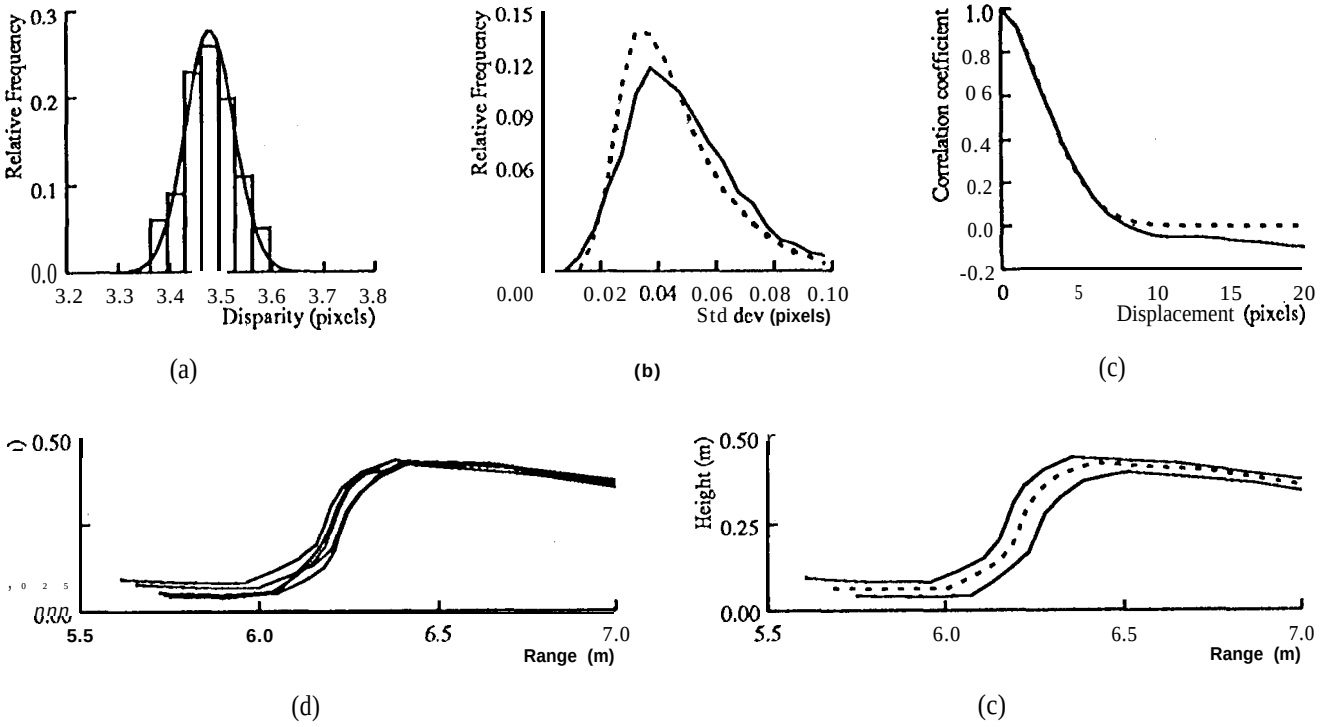


Figure 5: Dispersion statistics, (a) Histogram of disparity estimates at a single pixel, overlaid with Gaussian distribution having same standard deviation as the sample data. (b) Frequency plot of sample standard deviations of disparity for all pixels-in the image, for 60x64 (solid) and 120x1 28 (dashed) resolutions. (c) Average correlation of disparity estimation errors for pixels in the same column, empirical values (solid) and fitted model (dotted) for 60 x 64 resolution. (d) and (e) Range profiles for 60x64 resolution, for 30cm obstacle at 20 feet: (d) overlay of profiles from five stereo pairs; (c) envelope and mean (dashed line) of range profiles from 100 stereo pairs.

3.3.2 Dispersion

Figure 5 shows results of experiments to evaluate the theoretical model of random errors. Wc began by examining the assumption that disparity estimates arc Gaussian distributed. Figure 5a shows a histogram of disparity estimates for a typical pixel from 100 stereo pairs of the gravel road. A Gaussian distribution with these parameters is overlaid on the histogram. The correspondence is good enough for our present purposes.

To evaluate the overall precision of disparity estimates, wc computed the sample standard deviation of disparity estimates at each pixel, averaged over all images in the data set. Relative frequency plots of the standard deviations for disparity maps estimated at 60 x 64 and 120 x 128 resolution arc shown in Figure 5b. The average of the standard deviations in both cases is around 0.05, or 1/20th of a pixel. Because the signal to noise ratio of the images at the two resolutions may differ, it was not clear *a priori* how the precisions would compare; experimentally, wc see that doubling the resolution has doubled the effective precision (in terms of angular resolution). The average of the standard deviations is lower at the obstacle edges, around 0.03 pixels: this is duc to the higher contrast of the obstacle boundary as compared with the empty road (see Figure 2).

To evaluate the the spatial correlation of estimation errors, wc computed the sample covariance matrix

$$\widehat{cov}(y_1, y_2) = \frac{1}{N}\sum_{i=1}^{N}[\hat{d}(y_1) - \hat{\mu}_d(y_1)][\hat{d}(y_2) - \hat{\mu}_d(y_2)]$$

for an entire column of the disparity map, then estimated the sample correlation coefficient for successive displacements:

$$\hat{r}(\tau) = \hat{r}(y_1 - y_2) = \frac{\widehat{cov}(y_1, y_2)}{\hat{\sigma}_d(y_1)\hat{\sigma}_d(y_2)}$$

The results arc plotted in Figure 5c. As expected, disparity errors are very highly correlated for small displacements. The correlation drops to near zero at a displacement of eight pixels; since the matching window was 7 x 7, this conforms to the suggestion that overlap in matching windows is the dominant factor in correlating the errors, and is near zero when the matching windows do not overlap.

The dotted curve overlaid in the figure shows the function

$$r(r) \equiv \exp(-0.08\, r^{1.8}), \quad (5)$$

which matches the experimental data very well in the region of interest. We will incorporate this in the obstacle detectability model of the following section,

So far, we have examined random errors in disparity. It is still desirable to confirm the distribution models for range estimates obtained from disparity. In [7], we briefly experimentally examined the distribution of of range estimates and concluded that it conformed reasonably well with a Gaussian error propagation mode]. In this paper, we add a simple, experimental check of the spatial correlation already noted for disparity. Figure 5d shows segments of elevation profiles for five stereo pairs. The profiles tend to shift back and forth as a unit; that is, the inter-frame shift of the entire profile is much greater than the inter-pixel shift of pixels within the same frame. 'I'his confirms the correlated error model above.

3.4 Discussion

Based on these results, we draw the following conclusions about random and systematic errors in stereo range estimates. Regarding random errors, a model approximating errors as Gaussian at each pixel and spatially correlated over short distances is supported experimentally and appears to be adequate for use in deriving initial probabilistic models of obstacle detectability. For systematic errors, we have seen significant systematic errors in disparity estimation, such as the halo and sub-pixel effects noted earlier. It will be desirable to find algorithm refinements that reduce these effects; also, it may be necessary to take such effects into account when predicting resolution requirements for obstacle detection. Questions remain open about how well we can estimate the uncertainty of range estimates using contrast measures from a single image; however, we leave these in order to turn to modeling the quality of obstacle detection.

4 Quality of obstacle Detection

We will now use the models of range uncertainty obtained in the previous section to derive statistical models of the uncertainly in obstacle detection. As a starling point, we address a simple algorithm that we have used successfully in real-time demonstrations of obstacle detection on gravel roads [7], We believe that the basic techniques developed here will be applicable to more elaborate obstacle detection algorithms. The balance of this section parallels the structure of the previous section; we first review the obstacle detection algorithm, then derive statistical performance models for that algorithm, and then examine the validity of the models experimentally.

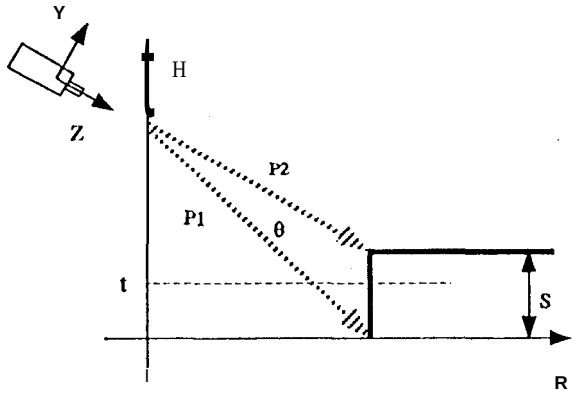


Figure 6: Obstacle detection algorithm

4.1 Obstacle Detection Algorithm

The obstacle detection algorithm is illustrated in Figure 6. Obstacles are assumed to be near-vertical step displacements on an otherwise flat ground plane. The algorithm checks for such obstacles by using pairs of pixels (p_1 and p_2) in the same column of the range image to measure the displacement in range and height between the two pixels. For each pixel p_1, the included angle between the p_1 and p_2 lines of sight is set to subtend a fixed obstacle size, denoted the *stepheight* S. Thus, the included angle θ is constant on each scanline, but varies from large at the bottom of the image to small at the top of the image. The 3-D vectors $\mathbf{P}_1$ and P_2 are estimated in the camera coordinate frame, then rotated into the vehicle coordinate frame and differenced to estimate the two components of surface slope. Starting with the disparities d_1 and d_2 measured at p_1 and p_2, respectively, the entire transformation is given by

$$\begin{aligned} \mathbf{P}_i &= \begin{bmatrix} Y \\ Z \end{bmatrix} = \begin{bmatrix} k_y y_i/d_i \\ k_z/d_i \end{bmatrix} \\ \Delta\mathbf{P} &= \begin{bmatrix} \Delta H \\ \Delta R \end{bmatrix} = \mathbf{R}(\mathbf{P}_2 - \mathbf{P}_1) \\ &= \mathbf{R}\left(\begin{bmatrix} k_y y_2/d_2 \\ k_z/d_2 \end{bmatrix} - \begin{bmatrix} k_y y_1/d_1 \\ k_z/d_1 \end{bmatrix} \right), \end{aligned} \quad (6)$$

where y_i is the image row coordinate of pixel p_i, $\mathbf{P}_i$ is a vector in camera coordinates, R is the rotation matrix that transforms between camera and vehicle coordinate frames, and k_y and k_z are constants that subsume camera calibration parameters for focal length, image aspect ratio, and stereo baseline. The obstacle detection decision rule is based on AP. Since the components of this vector are highly correlated, very little is lost by simply declaring an obstacle to exist at pixel p_1 if the measured change in height, AH, exceeds a given threshold t. As this also gives a simpler statistical model, in this paper we use only ΔH for clarity. As an example, to detect obstacles larger

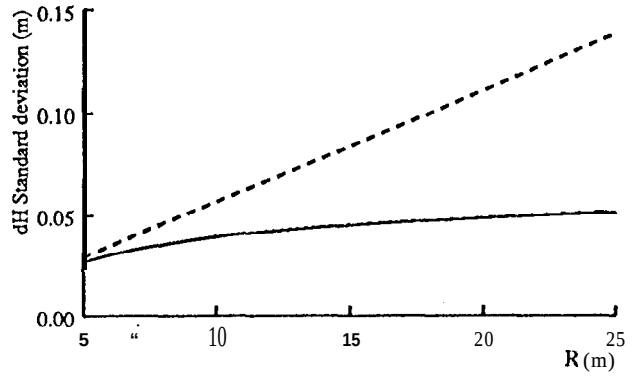


Figure 7: Theoretical standard deviation of measured change in height ($\sigma_{\Delta H}$) versus true range for 30 cm obstacle observed at 60 x 64 resolution, based on equation (7) with $\sigma_d = 0.05$ pixel. Solid curve.: $r(\tau) = \delta(0)$. Dashed curve: experimental $r(\tau)$ (equation (5)).

than, say 30 cm high, wc set S = 30 cm, measure ΔH, and decide there is an obstacle present if $\Delta H \geq t$, where $t < S$.

4.2 Statistical Performance Models

To derive a model of the reliability of obstacle detection requires a model of the probability density function (pdf) $f(\Delta H)$ of ΔH. To obtain this, wc will model $\Delta \mathbf{P}$ as jointly Gaussian and derive its covariance matrix $\Sigma_{\Delta \mathbf{P}}$ by linear error propagation from the covariance matrix Σ_{d_1,d_2} of d_1, d_2. Assuming that the uncertainly in disparity is the same at p_1 and p_2, this yields

$$\Sigma_{\Delta \mathbf{P}} = J\, \Sigma_{d_1,d_2}\, J^T = \begin{bmatrix} \sigma_d^2 J \begin{matrix} 1 & r(\tau) \\ r(\tau) & \end{matrix}, & J^T \end{bmatrix} \quad (7)$$

where J is the Jacobian of (6) and $r(\tau)$ is the spatial correlation coefficient of the estimation errors for d_1 and d_2. For obstacle detection based on thresholding *A}],* the reliability of detection will depend on the standard deviation $\sigma_{\Delta H}$ obtained through this error propagation. Figure 7 plots $\sigma_{\Delta H}$ versus range for an image resolution of 60 x 64 pixels, using experimental values of σ_d and $r(\tau)$ from the previous section. For this resolution, a 30 cm obstacle subtends about 6 pixels at 5 meters; therefore, ΔH estimates for this distance and greater will be affected by the correlated disparity measurements, as shown by comparing the solid and dashed curves in the Figure.

The next step in modeling the quality of obstacle detection is to derive the conditional probability of deciding that an obstacle occurs at pixel p_1, given that a step of height $\mathcal{H}$ actually exists there. This is obtained by integrating $f(\Delta H|\mathcal{H})$ over all possible measurements above the threshold:

$$p(\Delta H > t|\mathcal{H}) = \int_t^\infty f(\Delta H|\mathcal{H})\, d\Delta H$$
$$= \frac{1}{\sqrt{2\pi}\sigma_{\Delta H}} \int_t^\infty \exp\left(-\frac{1}{2}\frac{(\Delta H - \mathcal{H})^2}{\sigma_{\Delta H}^2}\right) d\Delta H$$

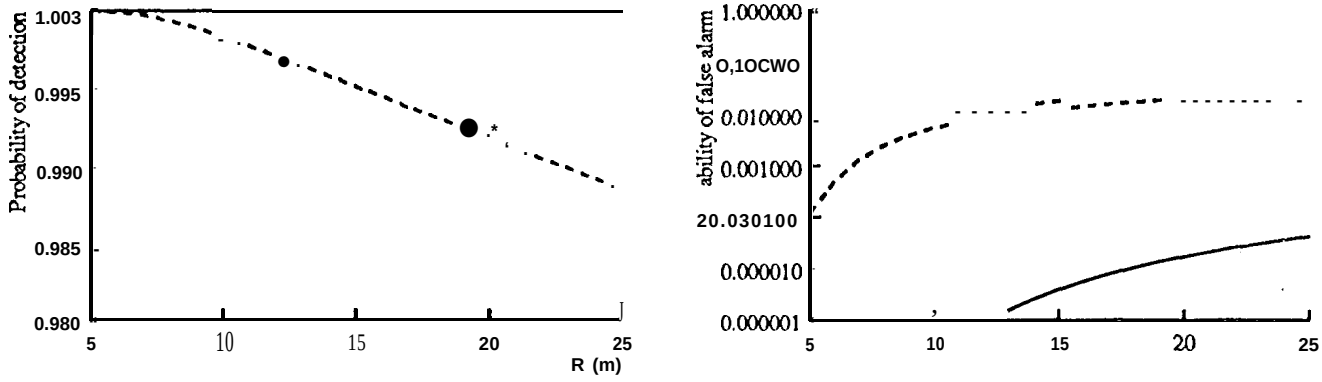
This conditional probability is a kcy tool for analyzing obstacle detection performance, because it embodies both the probability of detecting an obstacle that is actually present (P_d) and the probability of a false alarm (P_f) [8]. For example, the conditional probability P_d of detecting an obstacle of a size $\mathcal{H}_1$, given that such an obstacle is actually present (i.e. has its base at pixel p_1), is the integral $p(\Delta H > t|\mathcal{H}_1)$. Figure 8a shows P_d versus range for stepheight S = 30 cm, $\mathcal{H}$ = 30 cm, and a threshold t =20 cm. For a disparity standard deviation of σ_d = 0.03 pixels, which is the mean at obstacle edges, the model predicts almost perfect detection over distances from 5 10 25 m, For σ_d = 0.05 pixels, P_d decreases to 0.988 at 25 m. Since this model applies to detection performance at a single pixel, detection performance would be better for obstacles covering several pixels in width.

If no obstacle is present (e.g. $\mathcal{H}_0 \equiv \mathcal{H} = 0$), the probability of a false alarm is $P_f = p(\Delta H > t|\mathcal{H}_0)$. For illustration, Figure reff:od:pdpfvsrb shows P_f as a function of range. Again, this is a model for performance at a single pixel; therefore, estimating the number of false alarms in an image region requires integrating P_f over the area of the region. If detection performance were independent at each pixel, then the expected number of false alarms would be P_f multiplied by the number of pixels in the region. For 60 x 64 images, wc typically estimate disparity for on the order of 1000 pixels; using this as the region size, wc would need $P_f < 0.001$ to have less than one false alarm pcr frame. The model for σ_d = 0.05 pixels satisfies this for ranges up to 25 meters. Although this estimate dots not account for the spatial correlation of disparity errors, the experimental results below suggest that it is still a useful guide to the design or evaluation of the obtacle detector.

So far, our discussion of detection and false alarm probabilities has only considered simple hypotheses; that is, distinguishing between two particular alternatives $\mathcal{H}_0$ and $\mathcal{H}_1$. In practice, terrain height variations will occur over a continuous range of values. We have begun to address this issue in [10].

4.3 Experimental Evaluation

To evaluate the mode] for detection probabilities, wc computed detection statistics for the obstacle data sets of Figure 3. In this paper, wc focus the evaluation on one obstacle size at various distances. For false alarm probabilities, wc found that image ensembles of a still scene produced



(a) (b)

Figure 8: (a) Probability of detection (P_d) versus range for a 30 cm obstacle observed at 60 x 64 resolution, with stepheight parameter S = 30 cm and decision threshold t = 20 cm. Solid curve: σ_d = 0.03 pixels; dotted curve: σ_d = 0.05 pixels. (b) Probability of false alarm (P_f) versus range for S = 30 cm and t = 20 cm. Solid Curve: σ_d = 0.05 pixels; dotted curve: σ_d = 0.1 pixels.

results that were biased by the fixed, underlying intensity pattern. Therefore, wc generated an image ensemble with varying intensity signal by acquiring 50 images while driving a vehicle down the same flat, obstacle-free road. This has the added benefit that the results bear more directly on performance under actual driving conditions.

Figure 9 *shows* the sample mean and standard deviation of ΔH estimates at each pixel for the flat road sequence. The ripple pattern in the means reflects the sub-pixel biases discussed in section 3. For this data, the sample standard deviation of disparity σ_d averaged 0.13 pixels and the disparity estimation errors arc spatially correlated over a much larger distance than for the static data sets discussed on section 3. This can be explained by multiple causes, including the variations of the vehicle attitude when driving and the variations of the ground shape along the road. Nevertheless, using these noise parameters the trend in sample $\sigma_{\Delta H}$ values agrees well with our model.

To examine ΔH measurements with obstacles present, wc computed sample statistics of ΔH for image windows onc pixel high by three pixels wide, centered at the based of the left edge of the 30 cm obstacle. Figure 10a shows the sample mean of ΔH at each distance, with one-u error bars showing the sample standard deviation. For comparison, the figure also shows sample means and standard deviations for the same image windows in the flat road sequence. The means show that the obstacles were perceived as smaller than the true height of 30 cm. This is probably duc to the slant in the estimated range profiles (Figure 3). The standard deviations with obstacles present arc much smaller than without obstacles; this is duc to the smaller disparity noise at the obstacle boundaries (see section 3).

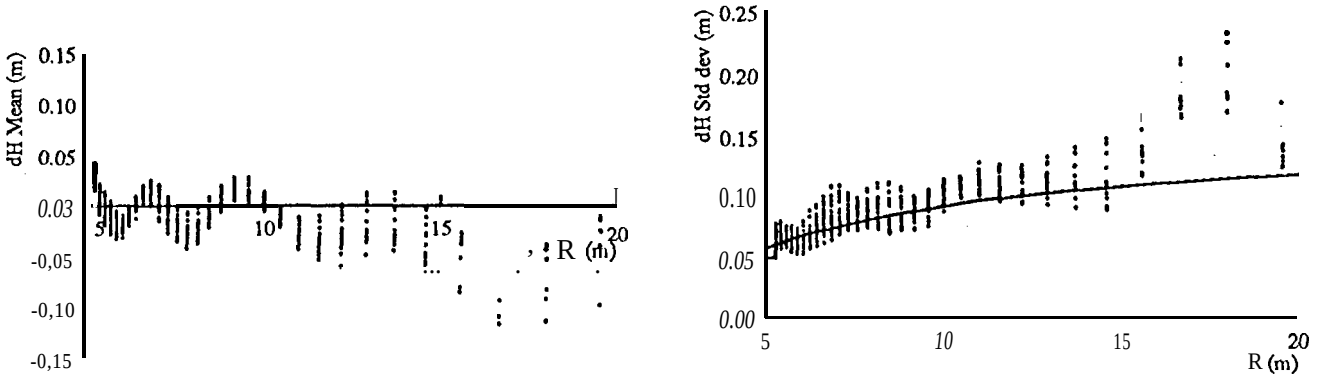
False alarm performance is shown in Figures 10b and 10c. The periodic variation, especially evident for a threshold of t = 15 cm, corresponds with the biased means of ΔH seen in Figure 9a. The theoretical P_f using the sample value of σ_d = 0.13 pixels provides an upper bound on the error rate that is fairly reasonable fort = 15 cm, but significantly the error rate for t = 21 cm. Reasons for this include some negative bias in the sample means (Figure 9a) and the fact that sample distributions of ΔH exhibit some skew.

For this data, a threshold of t = 21 cm gives perfect detection (P_d = 1) at all distances. Since the sample means of ΔH vary for each of the five distances, it would not be meaningful to plot P_d versus range for fixed t.

Figures 10d to 10f show detection and false alarm performance at 120 x 128 resolution. At this resolution, the sample standard deviation of disparity for the fiat road driving sequence is 0.2 pixels. The figures show that the uncertainty in ΔH and the probability of false alarms at each pixel arc lower at this resolution that at 60 x 64. Although it is not clear from the figure, the total number of false alarms pcr frame is also lower. At this resolution, with a threshold of 21 cm there arc no false alarms out to a distance of about 12 m. However, the false alarm rate over the distance range wc arc examining is now so low that the sample size needs to be increased to produce reliable sample statistics.

4.4 Discussion

As with the range data evaluation in section 3, for obstacle detectability wc find that high contrast intensity patterns and sub-pixel disparity errors induce notable biases in detection performance. Using ensembles of test images



(a) (b)

Figure 9: Empirical statistics of measured ΔH vs. true range at each pixel (as determined from nominal ground truth range data), for 60x64 resolution. (a) Mean at each pixel, (b) Standard deviation at each pixel, Overlaid curve shows theoretical $\sigma_{\Delta H}$ for disparity standard deviation σ_d = 0.13, which was approximately the observed value for this data.

that include different intensity patterns smooths out some of these effects; moreover, doing so make our evaluation conform more closely to real driving conditions.

For false alarms, the model for P_f provides an upper bound on error rates that is usable for predicting performance. Similarly, the model for P_d can bc used to predict detection rates, provided that wc also allow for biases in the mean of the perceived obstacle height, Together, these models contribute to answering the design questions posed in section 2; that is, they can be used to predict image resolutions that will achieve stated requirements on error rates in detecting obstacles of given sizes. Wc will conduct such analyses in future work.

5 Summary and Conclusions

Statistical modeling and evaluation of the performance of obstacle detection systems for UGV's is essential for the design, evaluation, and comparison of sensor systems. In this paper, wc addressed this issue for imaging range sensors by dividing the evaluation problem into two levels: quality of the range data itself and quality of the obstacle detection algorithms applied to the range data. Wc reviewed existing models of the quality of range data from stereo vision and AM-CW I. ADAR, then used these to derive a ncw model for the quality of a simple obstacle detection algorithm. This model predicts the probability of detecting obstacles and the probability of false alarms, as a function of the size and distance of the obstacle, the resolution of the sensor, and the level of noise in the range data. We evaluated these models experimentally using range data from stereo image pairs of a gravel road with known obstacles at several distances.

Wc draw two main conclusions from the results. First, the statistical modeling effort shows promise for being useful as a practical design tool, because it can be used to predict the reliability that will be achieved for specific resolutions or thresholds. Second, our experimental evaluation methodology gives valuable, quantitative measures of actual obstacle detection performance, independent of the predictive modeling issue, The experimental results also revealed artifacts in the range data that suggest areas for future algorithm research.

In the future, wc plan to refine the statistical performance modeling and evaluation work begun here and to usc it to (1) measure progress on stereo vision algorithms, (2) compare the performance of stereo vision with daylight cameras to other range image sensors, and (3) evaluate the performance of more elaborate obstacle detection algorithms. Also, wc believe that quantitative, statistical performance models like those presented here arc essential ingredients in algorithms for intelligent control of perception, or "active perception". Wc have begun to apply these performance models to perception control in [10].

References

[1] D. Nitzan, A. Brain, and R. Duda. The measurement and usc of registered reflectance and range data in scene analysis. *Proceedings of the IEEE*, 65(2):206–220, February 1977.

[2] C. C. (cd.) Slama. *Manual of Photogrammetry*, volume 4th cd. American Society of Photogrammetry, Falls Church, Vs., 1980.

[3] D. B. Gennery. *Modelling the environment of an exploring vehicle by means of stereo vision.* PhD thesis, Stanford University, June 1980.

[4] N. Ayache and O. D. Faugeras. Maintaining representations of the environment of a mobile robot. In *International Symposium of Robotics Research 4.* MIT Press, 1987.

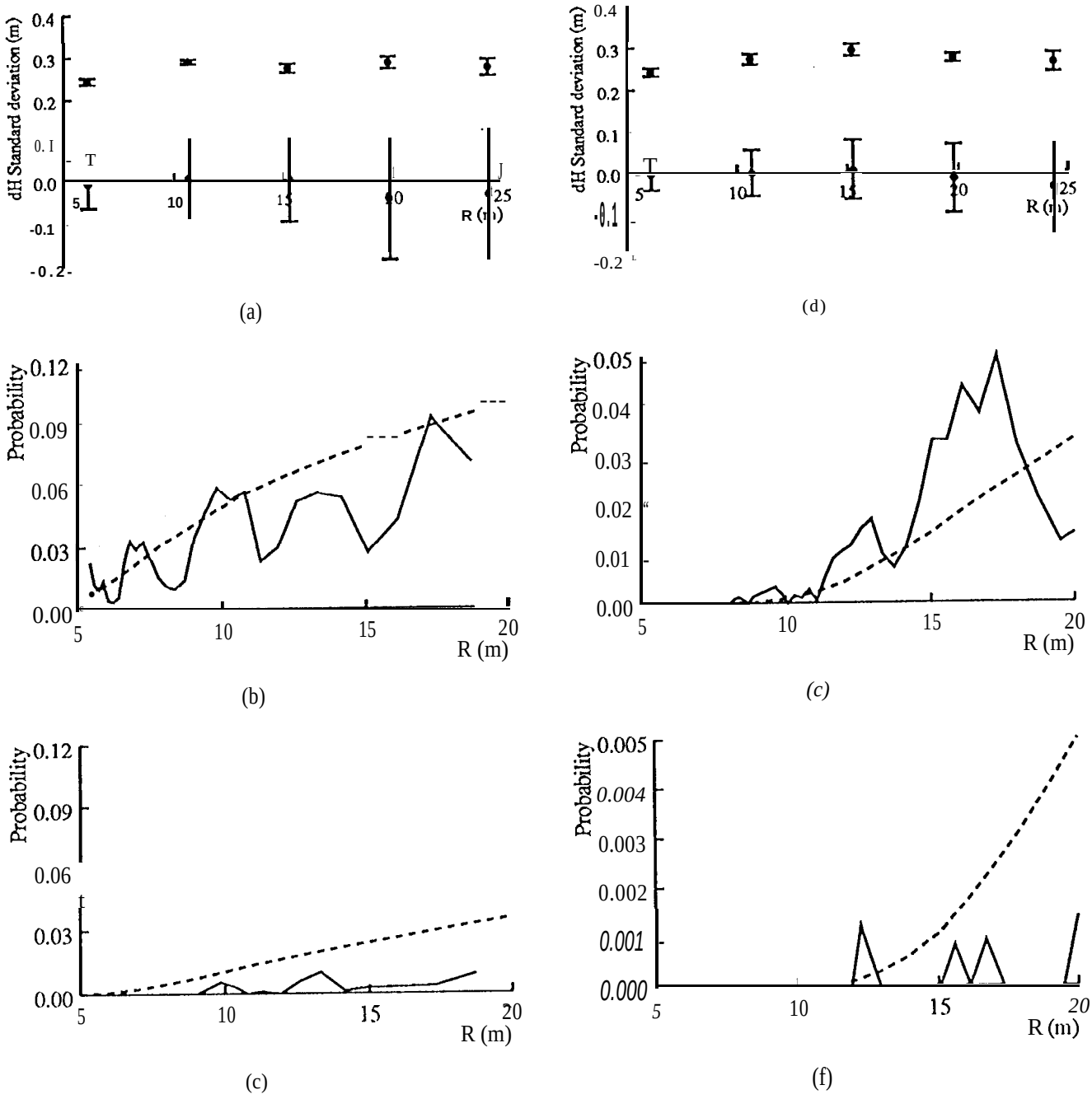


Figure 10: (a) *AH* mean at 60 x 64 resolution, S = 30 cm, for distances of 20, 35, 50, 65, and 80 feet. Top dots arc for 30 cm obstacle; bottom dots arc for flat ground. Error bars show ± onc standard deviation. (b) and (c) Probability of false alarms (P_f) versus range for thresholds oft = 15 and 21 cm, respectively. Solid curve: empirical results; dashed curve: theoretical model for σ_d = 0.13 pixel, (d), (c), and (f) Comparable results for image resolution of 120 x 128. Theoretical P_f curves arc for σ_d = 0.2 pixel.

[5] L. H. Matthies and S. A. Shafer. Error modeling in stereo navigation. *IEEE Journal of Robotics and Automation*, RA-3(3):239–248, June 1987.

[6] M. Hebert and E. Krotkov. 3D measurements from imaging laser radars: how good arc they? *Image and Vision Computing*, 10(3):170–178, April 1992.,

[7] L. H. Matthies. Toward stochastic modeling of obstacle detectability in passive stereo range imagery. In *Proc. Conf. on Computer Vision and Pattern Recognition.* IEEE Computer Society, June 1992.

[8] H. L. Van Trees. *Detection, Estimation, and Modulation Theory*, volume Part I. John Wiley and Sons, Ncw York, 1968.

[9] M. F. Reiley, D. C. Carmer, and W. F. Pent. 3-D laser radar simulation for autonomous spacecraft landing. In *Proc. SPIE Int'l Symposium on nigh Power Lasers.* SPIE, January 1991.

[10] P. Grandjean and L. Matthies. Perception control for obstacIC detection by a cross-country rover. Submitted to 1993 *IEEE International Conference on Robotics and Automation.*

11] L. H. Matthies. Stereo vision for planetary rovers: Stochastic modeling to near real-time implementation. *International Journal of Computer Vision, 8(1):71-91, 1992.*

12] P. S. Maybeck. Stochastic *Models, Estimation, and Control*, volume 1. Academic Press, New York, NY, 1979.

13] M. Okutomi and T. Kanade. A locally adaptive window for signal matching. Technical Report CMU-CS-90-178, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, October 1990.

14] A. K. Jain. *Fundamentals of digital image processing.* Prentice Hall, Englewood Cliffs, NJ, 1989.

[15] L. H. Matthies, R. Szeliski, and T. Kanade. Kalman filter-based algorithms for estimating depth from image sequences. *Interns/iotd Journal of Computer Vision*, 3(3):209–236, *1989.*

[16] M. H. DeGroot. *Probability and Statistics.* Addison Wesley, Reading, MA, 1975.