$$d\psi = \left(\frac{1}{R_{max}}\right)dx$$

The ideal row step varies greatly with range. It decreases quadratically with increased range.

$$d\theta = \left(\frac{h}{R^2}\right)dy$$

Luckily, however, it can be computed anew for each pixel because the range window is processed column by column. Although the flat world assumption may seem inappropriate on rough terrain, the use of it in adaptive scan works well in practice. Fundamentally, the most important factor affecting range pixel density is the local terrain gradient. If the range gradient is high in the image in the horizontal or vertical direction, then it is probably the case that the terrain is undersampled by the sensor, and the flat world assumption makes little difference because the data is of poor quality. Indeed, it may be more robust to ignore such data. If however, the gradient is small, the algorithm performs acceptably. A conceptual code fragment is given below:

```
/*
** Compute scan increments
*/
ds = map resolution
col_skip = ds/Rmax*image->cols/HFOV;
/*
** Process Range Window
*/
j = image->start_col;
while ( j <= image->end_col + col_skip)
   {
   i = image->end_row;
   while ( i >= image->start_row - row_skip)
      {
      if (range(i,j) > Rmax ) break;
      row_skip = ds*h/
(range(i,j)*range(i,j))*image->rows/VFOV;
      else if( range(i,j) < Rmin )
{i -= row_skip; continue;}
      else process_pixel_into_map();
      i -= row_skip;
      }
   j += col_skip;
   }
```

Figure 13 - Adaptive Sweep/Scan Ir

One of the virtues of the algorithm is that it intelligently moves toward the bottom of the range window by subsampling in the region of the image where the data density is extremely high. In actual implementation, oversampling factors are introduced into both the row and column skip factors as a safety margin.

13 Basics of Stereo Perception

For outdoor terrain, area-based stereo algorithms are typically used because it is necessary to estimate the range of every pixel in the image. The typical steps in the process are:

- Preprocess the images to enhance texture and remove bias and scale variations across the image. The output of this process is a normalized image which corresponds to each input image.
- For each candidate disparity considered, for a window around each pixel in the first image, compute a measure of correlation between it and a window around the pixel in the second image which is displaced by the disparity considered. The output of this process is a cube of numbers of the form Corr[i,j,d] which will be called the **correlation tensor**.
- The curve Corr[d] obtained by fixing the row and column indices of the correlation tensor will be called the **correlation curve**. For each pixel in the first image, the correlation curve is searched to find its extremum value. The value of the disparity at the extremum value of the correlation curve for that pixel is the quantity of interest. The output of this process is a **disparity image**.
- For each pixel in the disparity image, convert disparity to range using the stereo baseline. The output of this process is the **range image**.

9 Extra Leftovers

10 Computational Range Image Stabilization

This is directly due to the validity of the small incidence angle assumption.

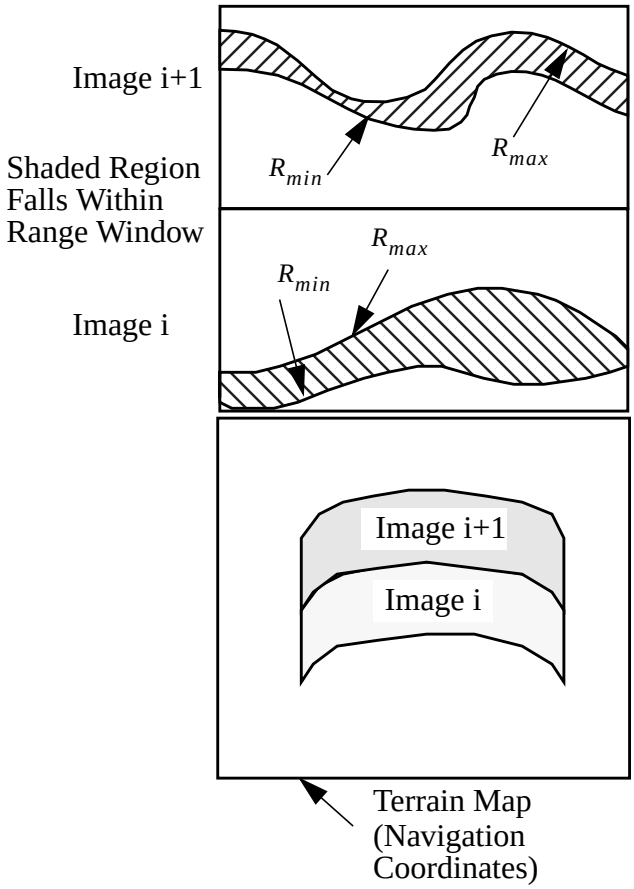


Figure 11 - Computational Image Stabilization

11 Range Window Computation - First Phase

A conceptual C code fragment is shown below:

```
/*
** Plan Window
*/
rhomin = minimum safe turn radius
treact = impulse turn reaction time (turn only)
speed = most recent speed estimate
range_offset = distance from vehicle frame to sensor frame
Pmax = speed * treact + rhomin;/* adaptive lookahead */
/*
** Range Window
*/
img_dens = imaging density
cycle_dist = distance travelled since last cycle
Pmin = Pmax - img_dens*cycle_dist; /* adaptive sweep */
/*
** Convert to image space - range feedbackward
*/
cur_dist = distance of most recent state estimate
img_dist = distance travelled when image was taken
delay = estimated latency of most recent state estimate
delay += system cycle time
fut_dist = cur_dist + speed*delay;
Rmax = Pmax + (fut_dist-img_dist)+range_offset;
Rmin = Pmin + (fut_dist-img_dist)+range_offset;
/*
** Add the wheelbase
*/
Rmax += wheelbase
Rmin += wheelbase
```

Figure 12: Adaptive Sweep Algorithm. The range window basis in order to robustly extract the data of interest.

```
j = start_col;
while ( j < end_col )
   {
   i = end_row;
```

12 Adaptive Scan in Range Images

However, the differential relationships are also a function of range. Consider attempting to keep dx and dy constant by varying the column step $d\psi$ and row step $d\theta$. The actual width of the range window is narrow in elevation angle, so the variation in ideal column step is small. Appropriate density is guaranteed by computing the column step at the maximum range because it is smallest here. Thus

is the ideal column step after it is converted to pixels.

disparity and range images. To the extreme right are the adaptively processed disparity and range images. The disparity images are shown to demonstrate the spurious matches which are caused by incorrectly chosen extrema in the correlation versus disparity curves.

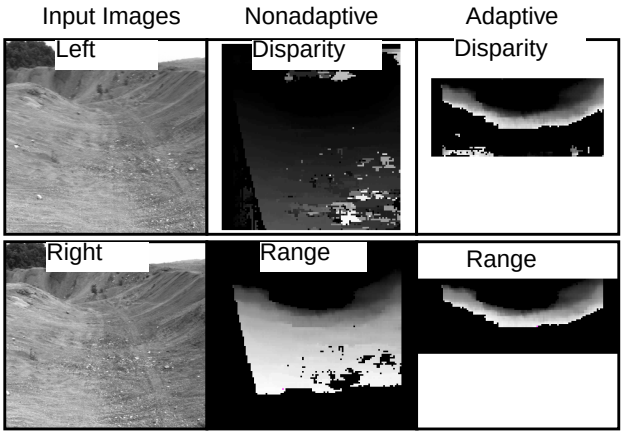


Figure 10: Adaptive Horizontal Baseline Stereo. The incorrect disparities due to incorrect matches are cleaned up with an efficient filter.

A breakdown of this run is shown in the table below:

Table 3: Stereo Adaptive Perception Performance (SPARC 20)

Attribute	Nonadaptive	Adaptive
Output Rows	120	48
Output Cols	128	128
Disparities	60	10
Preprocessing	102 msecs.	41 msecs.
Correlation	683 msecs.	69 msecs.
Postprocessing	754 msecs.	74 msecs.
Total Runtime	1539 msecs.	203 msecs.

7 Conclusions

It has been shown that an adaptive approach to perception based on the techniques of adaptive sweep and adaptive scan has several advantages. A more complete list of these advantages is as follows:

- It has the potential to solve or mitigate the perceptual throughput problem defined earlier.
- It computationally stabilizes a sensor within the limits of the sensor field of view. It converts an imaging sensor into an ideal adaptive scanner by adapting to vehicle speed, attitude and terrain shape.
- For limited field of view sensors, it provides an obvious basis for the generation of sensor pointing commands which keep a region of interest centered in the image.
- It solves the sampling problem for practical purposes because variation in the range ratio is very low over the small elevation width of the range window.

8 References

[1] R. Bajcsy, “Active and Exploratory Perception’, Computer Vision, Graphics, and Image Processing - Image Understanding, Vol 56, No 1, July 1992.

[2] M. Daily et al., “Autonomous Cross Country Navigation with the ALV”, In Proc. of the 1988 IEEE International Conference on Robotics and Automation, Philadelphia, Pa, April 1988; pp. 718-726.

[3] E.D. Dickmanns. “Dynamic computer vision for mobile robot control”, Proceedings of the 19th International Symposium and Exposition on Robots, pp. 314-27;

[4] R. T. Dunlay and D. G. Morgenthaler, “Obstacle Detection and Avoidance from Range Data”, In Proc. SPIE Mobile Robots Conference, Cambridge, Mass., 1986

[5] E. Grosso, and M. Tissarelli, “Active/Dynamic Stereo Vision”, IEEE Trans on Pattern Analysis and Machine Intelligence, Vol 17, No 9, pp. 868-879, Sept 1995.

[6] M. Hebert, T. Kanade, and I. Kweon. “3-D Vision Techniques for Autonomous Vehicles”, Technical Report CMU-RI-TR-88-12, The Robotics Institute, Carnegie Mellon University, 1988

[7] E. Huber, “Object Tracking with Stereo Vision”, Proc of Conf. on Intelligent Robots in Factory, Field, Space and Service. Houston, Tx, March 1994.

[8] A. J. Kelly, “An Intelligent Predictive Control Approach to the High-Speed, Cross Country Autonomous Navigation Problem”, Ph. D. thesis, Robotics Institute, Carnegie Mellon University, June 1995.

[9] In So Kweon, “Modelling Rugged Terrain by Mobile Robots with Multiple Sensors”, CMU PhD Thesis, 1990

[10] L. Matthies, “Stereo Vision for Planetary Rovers”, International Journal of Computer Vision, 8:1, 71-91, 1992.

[11] H. P. Moravec, “Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover”, Ph. D. Thesis, Stanford University, 1980.

[12] K. E. Olin and David Tseng. “Autonomous Cross Country Navigation”, IEEE Expert, August 1991.

[13] N. K. Papanikolopoulos, “The Framework of Controlled Active Vision”, Mathematical and Computer Modelling, Vol 25, No 5-6, pp. 145-163, Sept 1996.

[14] D. A. Reece and S. A. Shafer, “Control of Perceptual Attention in Robot Driving”, Artificial Intelligence, Vol 78, No 1-2, p 397-430, Oct 1995.

[15] M. J. Swain and M. A. Stricker, “Promising Directions in Active Vision”, International Journal of Computer Vision, Vol 11, No 2. Oct 1993.

[16] J. K. Tsotsos, “On the Relative Complexity of Active Versus Passive Visual Search”, International Journal of Computer Vision, Vol 7, Jan 1992.

range sensor and cameras used for the following results are given in the table below:

Table 1: Sensor Parameters

Attribute	ERIM laser rangefinder	CCD camera
Image Rows	64	640
Image Cols	256	486
Hor. Field of View	80°	20°
Vert. Field of View	30°	20°
Hor. Angular Resolution	0.3125°	0.0412°
Vert. Angular Resolution	0.4688°	0.0312°
Frame Rate	2 Hz	30 Hz

6.1 Range Image Adaptive Perception

In a typical image, the pixels that are actually processed by the adaptive perception algorithm form a horizontal band that is jagged-edged and of varying width. The width of the band decreases if the vehicle speed increases because adaptive lookahead will move the window up in the image where a smaller width projects onto the same groundplane distance.

The following figure gives a sequence of range images for a run of our navigation system simulator[2] on very rough terrain using a simulated rangefinder where the pixels that were actually processed fall between the thin black lines. On average, only 75 range pixels out of the available 10,000 (or 2%) were processed per image. In terms of areas imaged per second, the system throughput is increased by a factor of 100 times, or two orders of magnitude.

There are five range images arranged vertically on the left. These are rendered as intensity images where darker greys indicate increasing distance from the sensor. The terrain map constructed by the perception system is rendered on the right. The top figure shows the map as an image where lighter greys indicate higher elevations. In the center of the map is the vehicle at the position where the 5th image was captured. The lower right figure is the same terrain map rendered as a wireframe surface from the vantage point of the initial position.

There are three hills in the scene whose range shadows are clearly visible in the terrain map. In the first image, the vehicle is accelerating but still travelling relatively slowly. The range window is relatively wide and positioned near the bottom of the image. The first hill is in the range window. In the second image, the second hill is in the range window and the first hill has already been processed. Indeed, none of the left side of the image is processed because the data in the range window is occluded. In the third image, the third hill is now in the range window. In the fourth image, the vehicle is driving past the first hill and is rolled to the right because of it. This rolls the image to the left and the algorithm compensates appropriately. In the fifth image, the range window has moved past the third hill to the flats beyond and a fourth hill is barely visible in the distance.

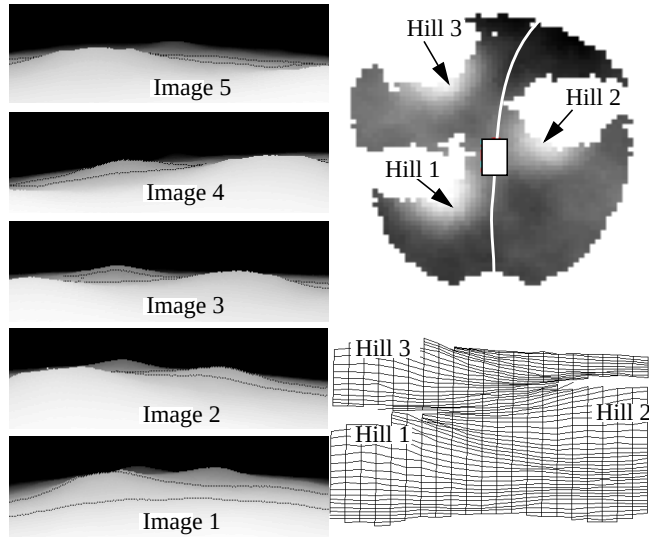


Figure 9: Adaptive Rangefinder Perception. The processing of five range images is illustrated as the vehicle drives through an obstacle course of three hills.

Actual perception performance is given in the tables below for a series of images of flat terrain. In the table, the nonadaptive value corresponds to the result obtained by processing all pixels in the ERIM range image. The adaptive value is the value obtained by our range image algorithm:

Table 2: Rangefinder Adaptive Perception Performance (SPARC 20)

Attribute	Nonadaptive	Adaptive
Pixels Processed Per Image	16384	75
Run Time	0.352 secs	0.022 secs

The results do not scale linearly with pixels processed because the adaptive result includes a constant setup time. Nonetheless, the adaptive result is 16 times faster than the nonadaptive result and if the ERIM sensor had higher angular resolution, the improvement would be proportionally better. The system uses barely adequate spatial resolution and eliminates redundant measurements and hence achieves minimum throughput.

6.2 Stereo Vision

The following figure illustrates the operation of embedded adaptive stereo on two horizontal baseline input images. These are images of a barren ravine road near CMU taken from inside the ravine. The initial input images appear at the left. To the right of these are the nonadaptively processed

[2]The system performs identically on real images but simulated ones were used here in order to illustrate several points in limited space.

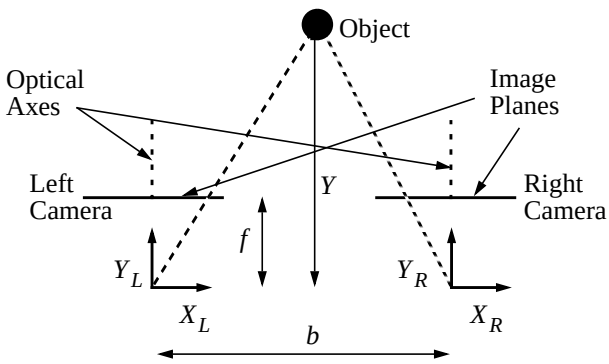


Figure 6: Stereo Triangulation. The relationship between disparity d, range Y, baseline b, and focal length f is derived from similiar triangles.

of view. In other words, the image coordinates of corresponding points in both images are very close to each other if the range of the point is beyond the response distance.

5.1.2 Local Minimum Problem

In traditional area-based stereo, correlations (or any of a number of other measures of similiarity of two image subwindows) are computed for a wide range of disparities. Then the algorithm searches along the curve generated for each pixel for the disparity, d^*, corresponding to the global correlation maximum. The case for normalized image crosscorrelation is illustrated below.

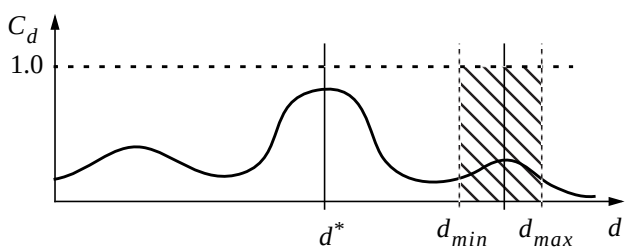


Figure 7: Disparity Search. The global maximum correlation is the best match. Limiting the search can lead to the wrong answer.

If, however, the search were limited to the disparity window whose boundaries are d_{max} and d_{min} in the above figure, the point of maximum correlation that would be found would only be a local minimum. No information other than the absolute value of the measure of similiarity would indicate this. If a range image were generated based on the results of this limited disparity search, the image would contain:

- correct ranges for pixels whose true range happened to fall within the range window searched.
- incorrect ranges for pixels like the one illustrated above which defeated our best attempts to identify them at this stage of processing.

Nevertheless, the environment is often smooth, and this smoothness leads to the property that correct ranges tend to form large smooth regions whereas incorrect ones do not as illustrated below.

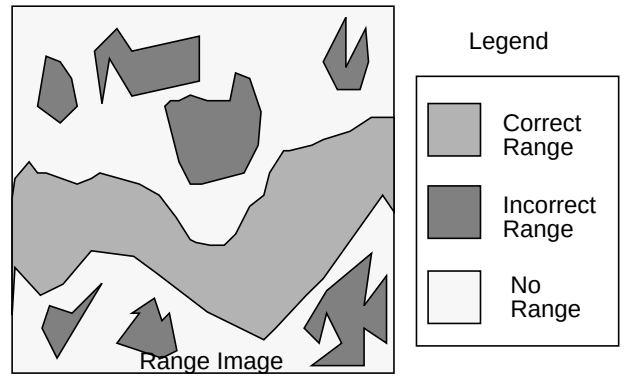


Figure 8: Spurious Disparities. Correctly ranged pixels tend to form large connected smooth regions. Incorrect ones do not.

It is well known that spurious matches occur fundamentally because regions which do not correspond physically actually look more or less the same. Several solutions to this repetitive texture problem help the situation somewhat but the simple technique of computing connected components and removing small regions [9] works effectively and is computationally free because a disparity image cleanup pass is required even when a wide disparity range is searched.

5.2 Embedded Adaptive Scan in Stereo Vision

In the case of stereo vision, the situation for adaptively changing resolution is more complex because range resolution and angular resolution are coupled. That is, once angular resolution is fixed, range resolution is also fixed, yet each has independent constraints imposed on it by the application. It is not possible, for instance, to aggressively reduce horizontal image resolution (as would be done with a range image) at the input to stereo because range resolution will also be dramatically and unacceptably degraded.

The least that can be done, however, is to compute the degree to which the output range image would be subsampled and then the latter stages of stereo (the stages past the correlation computation) can simply ignore the unwanted pixels. Before correlation, those unwanted pixels may be needed to participate in computing the correlations.

6 Results

The following two sections present performance results for adaptive perception based on laser range images and stereo vision. For these results, the vehicle speed is 3 meters/second and the resolution of the generated terrain map is 0.75 meters in both horizontal directions. An oversampling factor of 2 is also incorporated into adaptive scan as a safety margin to protect against terrain undersampling.

While adaptive perception resamples a range image for optimum coverage of the terrain, the specific attributes of the

responding to the constant image subsampling factors that give the most acceptable groundplane resolution. In the case of range images, adaptive scan is implemented by a literal subsampling of the image. Also, this subsampling applies to both the data in the ROI and the data below the ROI that is not processed. That is, adapting the resolution can benefit the speed of handling both the processed and the unprocessed data.

Because the differential transformation from the image plane to the groundplane is unknown, a perfectly robust, optimal subsampling solution is not available. However, a spectrum of approaches to resolution management are available based on the frequency of update of the row_skip and col_skip variables and how they vary with range for an assumed flat world. They can be computed based on:

- the highest projected value of the ROI maximum range, Rmax, based on the known speed limits of the vehicle.
- the value of ROI maximum range, Rmax, for the current computational cycle.
- the instantaneous value of range, R, at the current pixel.

These options have been listed in order of increasing speed and decreasing robustness.

In the least adaptive form of adaptive scan, the number of pixels skipped in the horizontal and vertical directions can be set based on the average or worst case expected value of the maximum range.

$$d\psi = \left(\frac{1}{R_{max}}\right)dx \qquad d\theta = \left(\frac{h}{R_{max}{}^2}\right)dy$$

In the next most adaptive form, the image plane resolutions are recomputed for each image based on the current ROI maximum range. In the most adaptive form, image plane resolutions can be recomputed based on the instantaneous range image values. However, it can be awkward to vary the azimuth resolution as a function of range if one chooses to process the image by columns.

The ratio of maximum to minimum range is normally small, so the variation in $d\psi$ (row_skip) is also small. Under this assumption, a good compromise is to use the worst case azimuth resolution and the instantaneously computed elevation resolution.

$$d\psi = \left(\frac{1}{R_{max}}\right)dx \qquad d\theta = \left(\frac{h}{R^2}\right)dy$$

Although the flat world assumption may seem inappropriate on rough terrain, the use of it in adaptive scan works well in practice.

5 Adaptive Sweep and Scan for Stereo Imagery

The principles of the earlier section could be applied directly to the output of a stereo vision system. Yet, because stereo also consumes computational resources, it seems worthwhile to investigate whether similiar techniques can be employed inside of the stereo algorithm itself in order to avoid computing range pixels that subsequently would be eliminated anyway.

Traditionally, the stereo problem is cast as one of determining the range for every pixel in the image. Traditional stereo finds the range for each possible angular pixel position. Conversely, our adaptive approach to stereo finds the angular positions in the image plane of each possible range value. It determines those pixels whose range value falls within a small range window, and it does so without computing the ranges of pixels which are not of interest. This principle is sometimes called **range gating** in laser rangefinders which employ it.

The motivation for the approach in the case of stereo is the observation that the region of terrain which is beyond the vehicle response distance usually corresponds to a very narrow range in stereo disparity space. The nonlinear relationship between range and disparity also implies that range resolution is relatively poor at high ranges, so the computation of the range of low range pixels can be wasteful. However, as before, the problem of selection, of determining membership in a range gate without computing the range, seems difficult.

5.1 Embedded Adaptive Sweep in Stereo Vision

For stereo ranging systems, the basic principle of the range window can be converted to a **disparity window**[1] for a stereo system because the range and disparity are related by the stereo baseline.

5.1.1 Disparity Window

The basic stereo configuration for perfectly aligned cameras is given below. It is useful to remove the dependence of disparity on the focal length by expressing disparity as an angle. Define the **normalized disparity** thus:

$$\delta = \frac{d}{f} = \frac{b}{Y}$$

Then, for a range window between 25 meters and 30 meters, and a stereo baseline of 1 meter, the angular width of the corresponding disparity window is:

$$\Delta\delta = \frac{1}{25} - \frac{1}{30} = 0.0067 = 0.38°$$

Thus, the range of disparities which corresponds to a typical range window is roughly 1% of a typical camera field

[1] There is a slight difference in the geometry of a stereo range image (perspective) compared to a rangefinder image (spherical polar . Therefore, a disparity window corresponds to a window on the y coordinate and not the true polar range. In most circumstances, this distinction can be safely ignored.

2.5.1 Minimum Computational Cost Implies Highest Speeds

The minumum computational cost of this approach to perception has implications for the real-time performance of autonomous vehicles. The maximum useful range of a perception sensor is often limited by reasons of eye safety, computational cost, limited angular resolution etc. Given this limit, the highest safe vehicle speeds are normally achieved by minimizing reaction times. The only element of reaction time that can be changed easily is often the component due to the time required to process imagery or perform other computations. Therefore, to the degree that our approach minimizes the computational cost of perception, it also increases the vehicle speeds that can be achieved.

2.5.2 Adaptive Sweep Implies Image Stabilization

Our software adaptive approach to perception has the side effect of computationally pointing the sensor vertical field of view by responding to both changes in the vehicle attitude and changes in the shape of the imaged terrain. While the shape of the range window may be very irregular in image space, it always corresponds to a regular semi-annulus in the ground plane. If the vertical field of view is wide enough and the range sensor is fast enough in terms of range pixel rate, this software adaptation is superior to the technique of physically stabilizing the sensor because it responds instantaneously.

3 Adaptive Lookahead for Range and Stereo Imagery

The three techniques described in the previous section can be applied to any range image generated by an imaging laser or radar sensor or a stereo vision system. It is also possible to embed adaptive perception into a stereo vision algorithm - which will be the subject of a special section. For both classes of imagery, range imagery and stereo pairs, the adaptive lookahead algorithm is common.

A vehicle may attempt to turn to avoid obstacles and maintain its forward speed, it may elect to stop completely, or it may choose any other arbitrary trajectory. The choice of trajectory determines the details of computing the response distance. For our purposes, adaptive lookahead is implemented by computing the distance required to execute a 90E turn at the current speed. This gives the maximum range of the range window.

The groundplane ROI must be defined very precisely in terms of distances from some specific point on the vehicle at some specific time. The problem of finding the data in this region in an image taken previously involves several aspects of time delays and geometric offsets.

- The sensor is not mounted at the vehicle reference point, so the ROI is adjusted for this offset.
- The vehicle is not itself a point, so the ROI must be enlarged to provide data at the positions of the wheels forward and aft of the reference point.
- There may be significant delay associated with the acquisition of an image, so the ROI must be adjusted for the age of the image.
- The most recent vehicle state estimate is itself somewhat old and computation takes finite time. The ROI may need to be adjusted for these effects depending on the instant with respect to which the ROI is defined.

4 Adaptive Sweep and Scan in Range Imagery

If one starts with a dense range image, the algorithm consists of the mapping of the range window into image space and the extraction of the data.

4.1 Adaptive Sweep

Terrain roughness and nonzero vehicle roll mean that the position of the range window in the image is different for each column so the range window is processed on a per column basis. In order to robustly find the range window, each column is processed in the bottom-to-top direction.

A conceptual C code fragment is as follows. The image itself is of dimensions rows by cols. A constant rectangular subwindow of the image is searched which is delimited by the image plane coordinates start_row, start_col, end_row, and end_col. This region is known to always contain the ROI but it is NOT responsible for most of the improvement in efficiency.

```
j = start_col;
while ( j <= end_col+col_skip )
   {
   i = end_row;
   while ( i >= start_row-row_skip )
      {
      R = range(i,j);
      if (R > Rmax )
         break;
      else if( R < Rmin )
         {i -= row_skip; continue;}
      else process_pixel_into_map();
         i -= row_skip;
      }
   j += col_skip;
   }
```

Figure 5: Adaptive Sweep Algorithm. The range window is processed on a per column basis in order to robustly extract the data of interest.

The **monotone range assumption** appears as the break statement after the first conditional of the inner loop. The start_col and end_col variables implement a fixed azimuth and elevation angle window within which the range window always lies on typical terrain.

4.2 Adaptive Scan

The variables row_skip and col_skip have values cor-

of the following approximations:

$$\frac{h}{R} \ll 1 \qquad \text{and} \qquad Y \approx R$$

We will call R the **range** and Y the **range projection**. It is easy to show that the relative error incurred in assuming that these two quantities are the same is the square of the ratio h/R. We will concentrate now on a specific class of region of interest - one that can be specified in terms of two range extremes. Let us define a **range window** or **range gate** as an interval given by R_{max} and R_{min} or, equivalently, by its corresponding range projection extremes Y_{max} and Y_{min}.

Suppose a pixel whose range projection is y is in a range projection gate:

$$Y_{min} < y < Y_{max}$$

Then, to first order, we have, under our assumption:

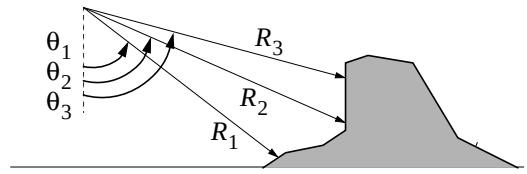
$$R_{min} < y < R_{max}$$

Which is to say that we can directly compare range pixel values (an image plane ROI) to a region of interest on the groundplane (a groundplane ROI) while incurring very little relative error. The small incidence angle assumption allows us to efficiently implement a test in the image plane of membership in a groundplane ROI. Under our assumption, it is not necessary to convert range pixel coordinates so it inexpensively decouples the problem of selection from that of perception. Only those pixels which satisfy the inexpensive image plane ROI membership test need have their coordinates converted.

2.3.3 Near Monotone Range Assumption

At this point, we have an efficient test for membership in a groundplane ROI. However, it is still expensive to test every pixel in a range image against a range gate. A final important assumption is the assumption that the environment is 2-1/2 dimensional with respect to the direction of gravity. That is, at all points, a line aligned with gravity pierces the first reflecting surface of the environment at most once. This assumption justifies a terrain map representation and it also allows us to assume that range is a near monotonic function of image elevation angle. The worst case violation of this **monotone range assumption** is the reduction in range that occurs when a vertical surface is scanned as shown below.

The computational advantage of the assumption is that once the maximum range is found in an image, all pixels above it in the same column of the image can be safely assumed to be beyond that range. It will turn out later that this assumption will only be used in laser rangefinder implementations of adaptive perception. Stereo vision will not require it.



$\theta_2 > \theta_1$ and $R_2 > R_1$ $\qquad$ $\theta_3 > \theta_2$ but $R_3 < R_2$

Figure 4: Nonmonotone range. Range is a nonmonotone function of image elevation angle at a vertical or near vertical surface.

2.4 Design

Our adaptive perception algorithm confines the processing of range geometry in any cycle of computations to an image plane ROI with the following properties:

- It extends beyond the vehicle response distance.
- Its size is the distance moved since the last cycle.

The algorithm has three conceptual parts as outlined below.

2.4.1 Adaptive Lookahead

Adaptive lookahead means the process of adapting the position of the groundplane ROI to assure that there is sufficient time to react to hazards. There is some minimum range inside of which it is unnecessary to look because the vehicle is already committed to travel there. Also, there is some maximum range beyond which it is unnecessary to look because there will be time to look there later. In detail implementation, the algorithm can set the minimum range to the response distance, or alternately, set the maximum range to response distance plus the distance travelled per cycle.

2.4.2 Adaptive Sweep

Adaptive sweep is the process of adapting the width of the groundplane ROI to assure that there are no holes or execssive overlaps in the coverage of the sensor. The ROI width is set to the distance travelled since the last computational cycle. This determines both the maximum and minumum range projections in the groundplane and they are trivially converted to the image plane ROI based on assumptions mentioned earlier.

2.4.3 Adaptive Scan

Adaptive scan is the process of managing resolution within the image plane ROI in order to achieve uniform groundplane resolution. For the data of interest, it will be possible to compute an approximate mapping from groundplane resolution to image plane resolution and images will be subsampled by appropriate factors to achieve near uniform groundplane resolution.

2.5 Implications

Certain implications of using the adaptive perception algorithm are worth noting here.

it robustly, but not both. This is the response-resolution tradeoff.

We will manage this tradeoff by explicitly computing the minimum distance required for robust obstacle avoidance and looking for obstacles only beyond this distance. This technique will be called **adaptive lookahead**.

2.2.2 Selection Problem

The mapping from the groundplane ROI to the image plane ROI is both nonlinear (a projective transform) and a function of the unknown shape of the terrain. It seems, therefore, that it is not at all straightforward to efficiently find the image plane ROI. Consider, for example, the straightforward solution of converting coordinates of all pixels in the image and then comparing their positions to the groundplane ROI. After pixels that are not in the groundplane ROI are eliminated, one is left with the image plane ROI. While this would certainly work, it can be far too inefficient to be useful.

The largest computational cost of a range pixel is the conversion of its coordinates from the image plane to the ground plane. In attempting to select only the data of interest by converting the coordinates of all pixels, one has already done most of the perception task anyway. Any straightforward attempt to selectively process data in a region of interest apparently falters because the problem of selection is as difficult as the problem of perception.

We will use assumptions to decouple these problems. When the assumptions are combined with an appropriate choice of the groundplane ROI, we will be able to partially infer the shape of the image plane ROI and compute its position by very efficient image plane search. The algorithm for doing this will be called **adaptive sweep**.

2.2.3 Sampling Problem

The **sampling problem** is the nonuniform and anisotropic distribution of pixels on the groundplane which corresponds to a uniform and isotropic distribution of the corresponding pixels in the image plane. The Jacobian matrix which relates the two distributions depends on both the image projective transform and the local terrain slope at each point. The impact of this problem is that not only is the shape of the image plane ROI distorted and of unknown position but the local pixel density required to sample the groundplane uniformly is both unknown and different everywhere in the image plane ROI.

This variation in pixel density is shown below for flat terrain. Each ellipse represents the footprint of a pixel. It is the variation in density which we are illustrating, not the density itself, so the images were subsampled to avoid clutter.

We will solve this problem to some degree by choosing the best compromize and, at other times, by actively computing the required image plane resolution from extrapolation. The algorithm for doing this will be called **adaptive scan**.

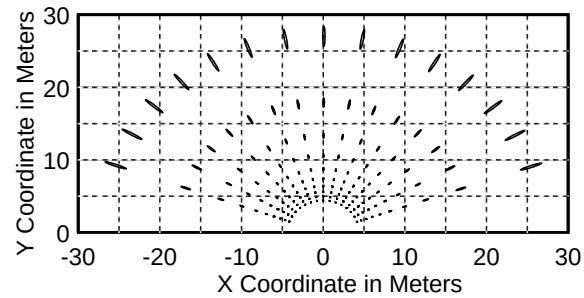


Figure 2: Sampling Problem. Equally spaced image pixels are not equally spaced on the groundplane - even for flat terrain. The situation is worse for rough terrain.

2.3 Assumptions

Certain assumptions will be key components of our approach - either because they must be made or because they can be made with little or no loss of generality.

2.3.1 Stationary World

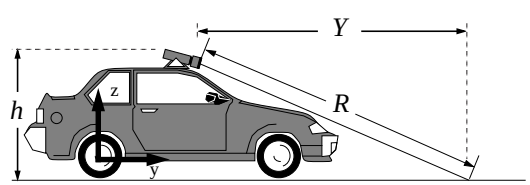
One of our most fundamental assumptions will be that the environment is self stationary. That is, the environment will be supposed to consist of rigid bodies whose relative positions are fixed - at least while they are in the field of view of the environmental sensor. While the bodies comprizing the environment are self stationary, our vehicle is in motion with respect to them. The value of this assumption is that it allows us to image a point in the environment only once and, because only the vehicle moves, its subsequent position relative to the vehicle at any later time can be inferred solely from the vehicle motion.

2.3.2 Small Incidence Angle

We will use the term **small incidence angle assumption** to refer to the situation where image pixels intersect a theoretical flat world at glancing angles. This is guaranteed to be the case if:

- the sensor is mounted on the vehicle roof, and
- pixels inside the response distance are ignored, and
- the vehicle speed is relatively high

because, under these conditions, the sensor height is small relative to the range of any pixel.



$$Y^2 + h^2 = R^2$$

Figure 3: Imaging Geometry. The height of the sensor above the ground plane is normally small compared to the ranges measured.

In the figure above, this assumption implies the validity

1.2 Perceptual Inefficiency

Luckily, analysis suggests [7] that much of the computational resources used to image and interpret the environment can be a waste of resources in mobility scenarios. This waste occurs for three principle reasons:

- Obstacles and other hazards normally appear in the field of view long before they can be resolved, and long after they cannot be avoided.
- The sensor vertical field of view is normally aligned with the direction of travel so that image sequences normally contain much redundant information.
- The projection of image pixels on the groundplane is normally elongated in the wrong direction for robust obstacle detection and minimum throughput.

We will show how to eliminate much of this inefficiency in order to generate perceptual throughput requirements that can be met easily.

One approach to reducing redundant information is the use of laser and video line scanners. These have seen use in specialized high-speed inspection applications for some time. In satellite applications, synthetic aperture radar has used vehicle motion to provide the scanning motion of the sensor along the direction of travel. The essential principle involved in these examples is to avoid scanning the sensor when either the motion of the vehicle or the motion of the environment already accomplish the scanning.

The approach of using line-scanned sensors suffers on rough terrain because abrupt attitude changes of the vehicle body cause holes in the coverage of the sensor. Software adaptation provides the best of both worlds because it gives the ideally focussed attention necessary for high speed and the wide field of view necessary for rough terrain.

2 Preliminaries

We will use two primary techniques for reduction of the perceptual inefficiencies mentioned above:

- We will actively maintain a **focus of attention** and process perceptual data only in a **region of interest** that contains the most useful information.
- We will actively and **intelligently subsample** the data within that region of interest for adequate - but not unnecessarily high - resolving power.

These two strategies will be referred to collectively as **adaptive perception** - the organizing principle of our approach to terrain mapping for high speed mobility.

2.1 Terminology

We will call a region of space for which sensory data is required a **region of interest**, abbreviated ROI.

2.1.1 Regions of Interest

It also will be important to distinguish the coordinate system implied by the sensor image - called the **image plane** from a set of coordinates attached to the terrain - called the **ground plane**.

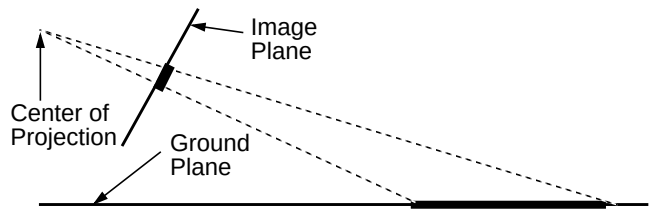


Figure 1: Regions of Interest. A region of interest in the ground plane forms a corresponding image in the image plane.

An ROI defined on the groundplane will be called a **ground plane ROI**. Such a region will have an image in the image plane which will be called an **image plane ROI**.

2.1.2 Differential Relationships

Let θ and ψ be the elevation and azimuth coordinates of an image pixel. Computing derivatives of the range image of flat terrain leads to the differential relationships between groundplane (x,y) resolution and image plane resolution (θ, ψ):

$$dx = Rd\psi \qquad\qquad dy = (R^2/h)d\theta$$

The completely correct transformations also depend on the local terrain gradients. These are unknown a priori because terrain geometry is the very thing the sensor is supposed to measure.

2.1.3 Response Distance

A quantity of central concern to us will be the distance that the vehicle requires to react to an external event such as the appearance of an obstacle in the sensor field of view. This distance will be called the **response distance** and its precise value will depend on:

- the speed of the vehicle when the event happens
- when the response is considered to be complete
- the maneuver chosen as the response

2.2 Subproblems

We have, at this point, nominated **adaptive perception** as a solution to the perceptual throughput problem. Unfortunately, this leads to a new set of problems, but we will be able to solve them with additional strategies and clearly identified assumptions.

2.2.1 Response - Resolution Tradeoff Problem

From the point of view of responding robustly to obstacles, it is best to detect obstacles early, or equivalently, at high range from the vehicle. However, from the point of view of sensor resolving power, it is best to detect obstacles as close as possible to the vehicle where data quality and spatial resolution tends to be highest. In other words, the farther away an obstacle is detected, the easier it is to avoid, but the harder it is to detect it robustly. When either resolution or range is limited, we can detect an obstacle robustly or avoid

Minimum Throughput Adaptive Perception for High Speed Mobility

Alonzo Kelly and Anthony Stentz

Field Robotics Center

Robotics Institute

Carnegie Mellon University

Pittsburgh, PA 15213-3890

email: alonzo@ri.cmu.edu, url: http://www.frc.ri.cmu.edu/~alonzo

email: axs@ri.cmu.edu, url: http://www.frc.ri.cmu.edu/~axs

Abstract

For autonomously navigating vehicles, the automatic generation of dense geometric models of the environment is a computationally expensive process. Yet, analysis suggests that some approaches to mapping the environment in mobility scenarios can waste significant computational resources. This paper proposes a relatively simple method of approaching the minimum required perceptual throughput in a terrain mapping system, and hence the fastest possible update of the environmental model. We accomplish this by exploiting the constraints of typical mobility scenarios. The technique proposed will be applicable to any application that models the environment with a terrain map or other 2-1/2 D representation.

1 Introduction

In this paper, we address one of the typical speed limitations of autonomous outdoor vehicles of the present generation - perceptual throughput. This quantity can be expressed in units of range or intensity pixels measured per second, or its equivalent. From the days of the Stanford Cart [11] to the Autonomous Land Vehicle [2], vehicle speed has been limited, at least in part, by limited perceptual throughput [6][9][8].

In this paper, we will exploit several assumptions that are valid in most outdoor mobility scenarios in order to selectively process only the data that matters in range imagery. In doing so, we will achieve near minimum perceptual throughput and hence, near maximum safe vehicle speeds.

Selective processing of visual data is not a new idea. This work falls into the research area of active or selective vision [1][15]. Active perception systems tightly couple perceptual components to the tasks that they support. They control aspects of the perception sensors or software in response to the needs of the task or to external stimuli. Previous work in active perception can be classified in terms of tracking, servoing, and homing systems [13][7], driving [14], sensor planning and object search [16], and mapping [5].

Our work falls into the mapping category though elements of all active vision applications are common. We will concentrate on selective data processing rather than sensor control and we will introduce several useful refinements to the obvious idea of continually looking somewhere new. In particular, by incorporating knowledge of vehicle motion and using justified geometric assumptions we are able to use fairly trivial image space search to massively reduce the computation required to map the environment.

The technique we use to selectively process data will be known as **adaptive perception** because it will adapt to changes in the speed of the vehicle, the attitude of the vehicle, and the slope of the terrain.

1.1 Terrain Mapping

When attempting to navigate over rough terrain, few assumptions about the shape of the terrain ahead can be made. It can be necessary to convert images into a full description of the geometry of the scene at relatively high rates. As a result, the speed of rough terrain navigation is typically limited by the throughput of the perception system. We will call this predicament the perceptual **throughput problem**.

For autonomous navigation purposes, natural outdoor terrain is well approximated by a surface expression of the form $z = f(x, y)$ where the z axis is aligned with the local gravity vector and the x and y axes are parallel to the local earth tangent plane. An important exception to this assumption is trees and other large vegetation which overhang their own supports. We assume either that the terrain is barren or that we can safely fill in the space beneath the branches in our models.

The surface of surrounding terrain can be sensed by any number of means, but the two most common ones are laser rangefinders and stereo vision. We represent the surface of the surrounding terrain by a sampled, uniform density data structure called a **terrain map** or **cartesian elevation map**.