Adaptive Perception for Autonomous Vehicles

Alonzo Kelly

CMU-RI-TR-94-18

The Robotics Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213

May 2, 1994

This research was sponsored by ARPA under contracts "Perception for Outdoor Navigation" (contract number DACA76-89-C-0014, monitored by the US Army Topographic Engineering Center) and "Unmanned Ground Vehicle System" (contract number DAAE07-90-C-RO59, monitored by TACOM).

Abstract

A new approach to autonomous vehicle perception is presented which solves the historically significant throughput problem at contemporary speeds through computational stabilization of the sensor sweep. This adaptive approach to perception has made it possible to achieve unprecedented autonomous vehicle speeds at little or no cost to other aspects of performance.

In order to measure the local environment at sufficient resolution and sufficient rate, an autonomous vehicle requires computational throughput on the order of $O[TV^2]$ where T is the vehicle reaction time and V is the velocity. On the other hand, the traditional approach of nonadaptive range image processing requires throughput on the order of $O[T^4V^4]$. The product TV is on the order of 10 for a conventional automobile so the difference between these two expressions is four orders of magnitude at 20 mph. Nonadaptive range image processing requires about 1 gigaflop in order to achieve 20 mph speeds whereas the algorithm presented here requires 1/10 of 1 megaflop under identical assumptions.

This report concentrates on the adaptive perception algorithm which forms the basis of RANGER's Map Manager object. The techniques used should be applicable to any application that models the environment with a terrain map.

Table of Contents

List of Figures

1. Introduction

One of the distinguishing characteristics of high speed autonomy is that the ratio of sensor height to vehicle stopping distance is small and, as a result, range pixels normally intersect the terrain at glancing angles. On the surface, this causes many problems, but finer analysis indicates that a consistent application of the small incidence angle assumption leads to solutions to some of those same problems. A second aspect of the problem is that the vertical field of view of imaging sensors is normally aligned with the direction of travel and therefore image sequences normally contain massively redundant data. A direct management of data redundancy through surprisingly trivial adaptive perception techniques leads to theoretical near minimum perceptual throughput.

A simple elegant active perception algorithm is introduced, based on these assumptions, which computationally stabilizes the sensor sweep. It solves the historically significant throughput problem in practical terms, and is four orders of magnitude more efficient than straightforward complete processing of all images at 20 mph. In converting any range imaging sensor to an ideal adaptive line scanner, it obviates the need for parallel processing in perception and makes unprecedented vehicle speeds possible on general purpose computer hardware. The system achieves the throughput necessary for 20 mph rough terrain autonomy on a typical engineering workstation.

The key problem of range image perception is that the position of the end of a range pixel is unknown until it is computed, and the computation of its location is the largest element of the computational expense of a pixel. Any attempt to selectively process data in an area of interest falters because the problem of selection is as difficult as the problem of perception. Luckily, for high speed autonomy, there is a key assumption, the **small incidence angle assumption**, which allows the circle to be broken, and allows selective processing of image data.

Adaptive perception is implemented for laser rangefinder images by actively searching for the data that lies in a **range window** in the image. Adaptive perception is implemented for stereoscopy by noticing that a range window corresponds to a **disparity window**. Other techniques are available to improve matters even further.

RANGER is an acronym for Real-time Autonomous Navigator with a Geometric Engine. This report describes the adaptive approach to perception which allows the RANGER system to support unprecedented vehicle speeds while still guaranteeing vehicle safety.

1.1 Acknowledgments

None of this work would have been possible without the guidance of Martial Hebert. Martial taught me terrain mapping, laser rangefinders, and guided the development of adaptive stereo. Many conversations with Martial led me into the search for a different way of doing range image perception.

1.2 Commentary

The throughput problem of autonomous navigation has long been regarded as a fundamental limitation and many organizations continue research into parallel processing approaches to the problem. This view of the difficulty of the problem is a natural conclusion if we adopt the assumption that entire images need to be processed at high rate. There are situations where this assumption is valid, and others where it is not.

The image processing view is necessary when the environment is dynamic and unpredictable because then the vehicle cannot predict the motion of important environmental features. It has to look around for them. It is also necessary in order to track stationary or moving environmental features for its own sake or for position estimation purposes.

Another justification for the image processing view is the poor quality of images relative to the application requirements. In this case, estimation theory can be used to merge redundant measurements into an acceptable overall estimate of the state of the environment.

Under an assumption of a static environment, a moving vehicle can reasonably expect an obstacle to stay put, and to therefore "scroll by" as the vehicle itself moves. All other things being equal, an obstacle needs to be seen only once, provided the vehicle can track its own motion accurately enough, and provided the perception of the nature and location of the obstacle is of adequate fidelity.

Even if the environment is not self-stationary, there is a second justification for a selective approach to environmental perception at high speed. The capacity of the vehicle to react to significant external events becomes ever more challenged as speeds increase. On this basis, it is possible to waste most or all perception cycles trying to see obstacles that cannot be avoided anyway. Thus, the processing of perception data that corresponds to regions too close to the vehicle is fundamentally unsound for obstacle detection and avoidance purposes.

In extreme cases, this logic, coupled with the extreme density of image data near the vehicle, leads to the conclusion that most perception processing is completely useless at high speeds. Further, limited throughput is the practical justification for poor angular resolution - because high resolution data cannot be processed anyway. Therefore, the historical tradeoff has been to provide inadequate angular resolution at the high ranges where it is needed in order to provide high resolution data close to the vehicle - when it is too late to use it.

To accept the idea of adaptive perception is to also accept the new requirement to remember what was seen previously. This requirement for a "map" of some kind, justified on throughput grounds, can also be justified on a more fundamental level. The idea of selective, or active, or adaptive perception has been around for some time. It is clearest in the systems which track features from image to image based on predictor/corrector algorithms and small search windows that move between images. Here, this idea will be applied to the problem of terrain mapping in outdoor rough terrain.

2. Analytical Basis of the Concept

This section presents some of the key considerations which drive the design of the perception system. These are presented in more detail in [15].

2.1 Guaranteed Safety

In order to guarantee that the vehicle remains safe, the automatic control system must ensure that the following four requirements are met simultaneously for all time:

- **Guaranteed Response** - The system must ensure that important environmental events are perceived in time to respond accordingly.
- **Guaranteed Throughput** - The system must ensure that it never drives over unknown terrain.
- **Guaranteed Detection** - The system must be able to resolve the smallest feature of the environment that can present a hazard.
- **Guaranteed Localization** - The system must be able to locate hazards relative to the vehicle to sufficient accuracy.

2.2 Throughput Problem

The throughput required to process an image depends on the number of pixels in the image. The number of pixels depends on the field of view and resolution. Resolution depends on the acuity requirement which implies it depends on range. The response requirement implies that range depends on speed so that resolution depends on speed. The throughput requirement also implies that field of view depends on speed. So ultimately, throughput can be expressed solely in terms of speed.

With an analysis of response and acuity it is possible to analyze the computational complexity of perception. In intuitive terms, guaranteed response implies that throughput is proportional to a high power of velocity because:

- Maximum range increases quadratically with speed (because braking distance does)
- Pixel size increases quadratically with maximum range (in order to resolve obstacles)
- Throughput increases quadratically with pixel size (assuming fixed field of view)

This relationship is indicated in the following figure:

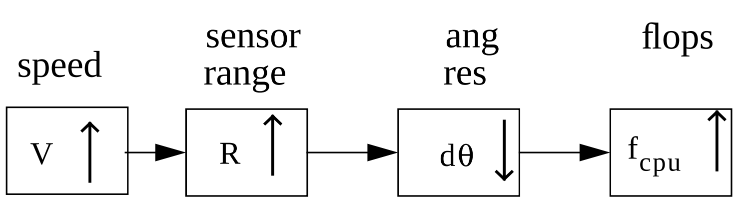


Figure 1 - Throughput Problem

When throughput is limited, this relationship gives rise to a trade-off between speed and resolution. Naive analysis suggests that the problem of high speed navigation is nearly impossible, because the necessary throughput is impractical. This will be called the **throughput problem**.

2.3 The Illusion

From an image processing perspective, the throughput problem appears to be impossible. Consider that contemporary rangefinders are 10 mrad resolution and many researchers believe that 1 mrad or so is needed to resolve obstacles for conventional automobiles. A ten fold increase in resolution is a hundredfold increase in pixels and a hundredfold increase in required throughput. Today, it is not possible to process 10 mrad images fast enough on a 10 Mflop processor. Therefore, if resolution were increased tenfold, it would be impossible to process 1 mrad images on a 1 Gflop processor. Brute force is not the elegant way to solve this problem.

On the other hand, the raw requirement is the throughput requirement, and this is trivial to meet. Consider that a 5 m/s vehicle covers about 6 map cells[1] between images at 2 Hz, so there is band in the image about six pixels wide[2] which would supply exactly the needed steady state throughput.

This section will show that ***the throughput problem is an illusion*** which arises from an image processing view of the problem and that simple adaptive techniques can solve it completely at contemporary sensor resolutions.

2.4 Adaptive Perception

It is possible to solve the throughput problem while simultaneously guaranteeing safety as efficiently as is theoretically possible by employing four principle mechanisms.

- **Adaptive Lookahead** is a mechanism for guaranteeing that the vehicle can respond to any hazards that it may encounter at any speed.
- **Adaptive Sweep** is a mechanism for guaranteeing barely adequate throughput and the fastest possible reaction time. In this way, speed is maximized.
- **Adaptive Scan** is a mechanism for ensuring barely adequate resolution that is as constant as possible over the field of view. In this way, speed is maximized without compromising robustness of the system.
- **Adaptive Regard** is a mechanism for ensuring that the system minimizes the spatial extent of the region it perceives based on vehicle maneuverability so that speed is maximized without compromising safety.

Together, these mechanisms can increase the efficiency (measured in terms of range pixel throughput) of a system by four orders of magnitude at 20 mph while simultaneously making it considerably more robust.

1. Assuming 1/2 meter cell resolution.
2. Assuming the image resolution is adequate to land one pixel in a map cell.

2.5 The Fundamental Speed/Resolution Trade-off

Identical resolution assumptions lead to the following throughput estimates for different image processing algorithms:

Table 1: Throughput Estimates

Algorithm	Estimate at Minimum Acuity, 4 second Reaction Time, and 10 m/s speed	Complexity
constant flux	250 Mflops	$O(T^4_{react} V^4)$
adaptive sweep	0.7 Mflops	$O(T^2_{react} V^3)$
adaptive sweep, scan	0.035 Mflops	$O(T_{react} V^2)$
ideal	0.0045 Mflops	$O(T_{react} V^2)$

The actual data for all 4 second reaction time curves is plotted below on a logarithmic vertical scale.

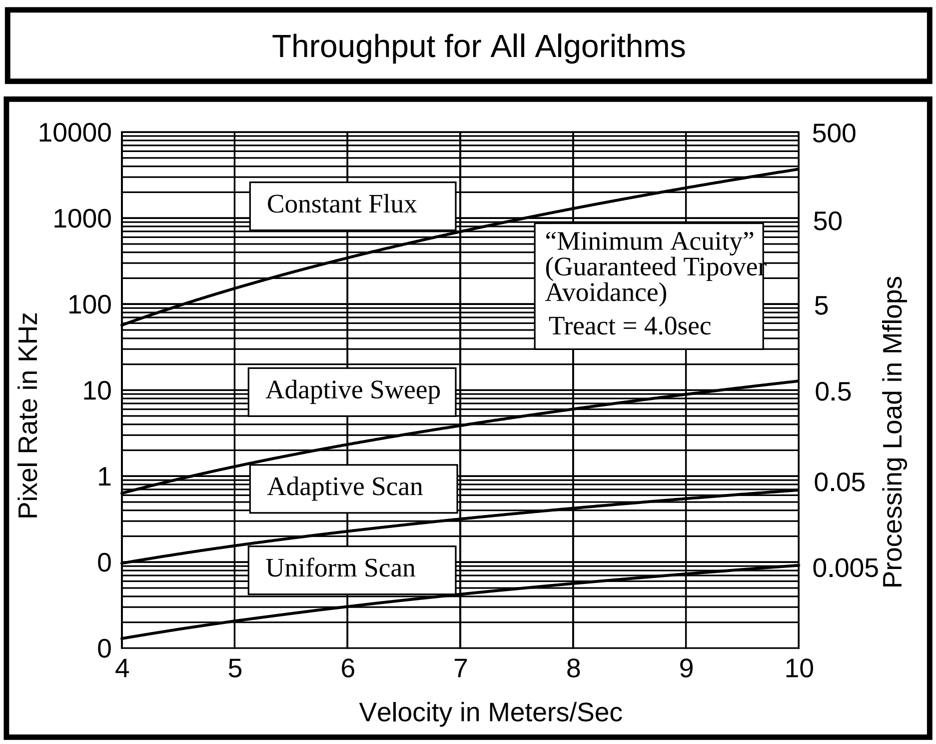


Figure 2 - Throughput for All Algorithms

The logic of decreasing pixel size for higher speeds is inescapable, but equivalent logic leads to a reduced vertical field of view requirement, so if the vertical field of view is not reduced, extreme throughput waste is being tolerated. Further, because pixel aspect ratio is extremely elongated at high ranges, the density of measurements in the crossrange direction is grossly suboptimal unless it is managed.

Notice that the complexity in either the above cases contains a constant times a power of the product $T_{react}V$. That is:

$$f_{cpu} \sim \frac{1}{\eta_S} O\left([T_{react}V]^N [V]^M \right)$$

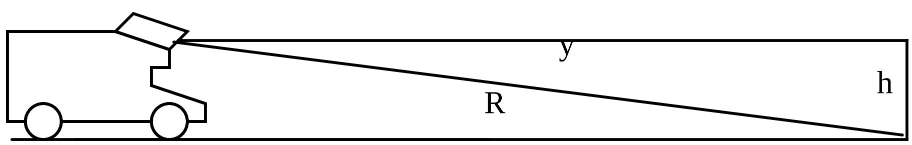
This will be called the **fundamental trade-off** because it indicates that the trade-off of finite computing resources is one of reliability for speed. This is a basic trade-off of speed and resolution which always arises from a system throughput limit. Computing resources establish a limit on vehicle performance which can be expressed as either high speed and low reliability or vice versa.

There are a few ways to read the result. If throughput is fixed, then speed is inversely proportional to reaction time. If speed is fixed, throughput required is the nth power of reaction time. If reaction time is fixed, throughput is the (n+m)th power of speed.

2.6 Small Incidence Angle Assumption

It turns out that the low **perception ratio** h/R which causes problems with respect to scanning density gives some payback with respect to adaptive regard and the sampling problem. On the surface, adaptive regard appears to be impossible because the mapping from image space to world space is unknown until it is computed. Once a pixel is computed, it might as well be used. This logic leads to processing the entire image and it is doomed to failure as was shown earlier. The only remaining question is how to do it.

Luckily, the range measurement from the sensor to the environment is almost identical to its groundplane projection (because the angle involved is so shallow). Indeed, the relative error in assuming the two are equal is the square of the perception ratio:



$$y = R\cos\theta \qquad \therefore \frac{y}{R} = \cos\theta \approx 1 - \left(\frac{h}{R}\right)^2 \qquad \therefore \frac{(y-R)}{R} \approx \left(\frac{h}{R}\right)^2$$

Figure 3 - Small Incidence Angle Assumption

This factor is on the order of 1% for high speeds simply because the range is so large. The assumption of a small perception ratio will be called the **small incidence angle assumption**. If the perception system attempts to process all geometry within a **range window**, the quest for the end of the range window will automatically walk right to the top of the image if necessary and, as a side effect, it will discover the height of a near vertical surface as long as it remains within the range window. This mechanism is far superior to simply processing a fixed subset of the vertical field of view (an **elevation angle window**) in part because of its performance on near vertical surfaces. Such surfaces would fall outside and elevation angle window and would not be completely processed.

The processing of pixels outside the range window is comprised of nothing more than reading their values and comparing them to the window. Further, because a terrain map already makes a 2-1/2 world assumption, the range values can be ***consistently*** assumed to be monotonic in elevation angle and this further reduces the required processing because some pixels need never be visited at all. This **monotone range assumption** also provides the basis for ambiguity removal in phase ambiguous sensors like AM rangefinders.

It was shown in [15] that as the minimum range increases, the scanning density decreases quadratically and approaches one. Therefore, because adaptive regard discards high depression scanlines anyway, ***the sampling problem is far less severe at high speed***.

3. Adaptive Perception From Range Images

In the case of laser rangefinders, a range image is generated in hardware. For this class of sensors, adaptive perception is limited to extraction of the range window from the complete image. The **adaptive perception** algorithm constitutes a simultaneous implementation of **adaptive lookahead**, **adaptive sweep**, and **adaptive scan** in one place. It proceeds in two phases; the computation of the range window based on the current speed and cycle time, and the mapping of this window into image space and the extraction of the data from the image while converting it back into world coordinates.

3.1 Range Window Computation - First Phase

A **range window** is computed which computationally points the sensor vertical field of view based on the guaranteed response and guaranteed throughput requirements. Adaptive lookahead is implemented by computing the distance required to execute an impulse turn[3] at the current speed. This gives the maximum range of the range window. Adaptive sweep is implemented by computing the distance travelled based on the cycle time and the speed and subtracting this distance from the maximum range. This adaptive sweep algorithm will automatically adapt to any system cycle time or sensor frame rate using small sweeps for faster sensors.

The true range window must be modified for three effects. First, the sensor is not mounted at the vehicle control point, so the above range window, the planning window, is adjusted for the offset of the sensor in order to project the window into sensor coordinates. Further, the vehicle is not itself a point, so adaptive regard is implemented by adding the vehicle wheelbase to the maximum range so that the largest possible hazard, a pitch hazard, can be detected in the detection zone beyond this distance. Third, there may be significant delay associated with the acquisition of an image, so the range measurements are adjusted for the age of the image. This is the true range window. A conceptual C code fragment is as shown below:

```
/*
** Plan Window
*/
Pmax = speed * treact + rhomin;/* adaptive lookahead */
Pmin = Pmax - speed * cycle_time /* adaptive sweep */
/*
** Range Window
*/
Rmax = Pmax + speed * sensor_latency - sensor_y + wheelbase;
Rmin = Pmin + speed * sensor_latency - sensor_y;
```

The fact that the last line does not subtract the wheelbase is a practical measure to ensure that all geometry used in planning comes from one image. This solves the image registration problem at a slight cost in redundant computation.

3. An impulse turn was chosen as the maneuver upon which to base adaptive regard because a vehicle which stops when it sees an obstacle (panic stop maneuver) is not useful when obstacles are dense. The impulse turn is a turn from zero curvature to the maximum. Other distinguished maneuvers are the turning stop and the curvature reverse.

3.2 Range Window Processing - Second Phase

In the second phase of adaptive perception, adaptive scan is implemented by simply skipping columns in the image. Up to 7 columns in 8 can be skipped for typical geometry and square pixels. This maintains guaranteed throughput and essentially uniform coverage. More sophisticated forms of adaptive scan were attempted, but the cost of computing the imaging Jacobian outweighed the gains achieved, so the simpler constant technique survives. It would be a simple matter to filter the input image in azimuth only, but it has not been implemented at this point, because it seems unnecessary on the ERIM sensor.

By the **small incidence angle assumption**, the projection of the sensor range onto the groundplane is essentially the groundplane y coordinate. However, terrain roughness and nonzero vehicle roll mean that the position of the range window in the image is ***different for each column***. Thus the range window is processed on a per column basis. A conceptual code fragment is as follows:

```
/*
** Process Range Window
*/
int i,j;
j = image->start_col;
while ( j < image->end_col )
  {
  i = image->end_row;
  while ( i > image->start_row )
    {
    if (range(i,j) > Rmax ) break;
    else if( range(i,j) < Rmin ) {i--; continue;}
    else process_pixel_into_map();
    i--;
    }
  j -= image->col_skip;
  }
```

The monotone range assumption appears as the break statement after the first conditional. The start_col and end_col etc. variables implement a fixed azimuth and elevation angle window within which the range window always lies on typical terrain. Ignoring the constant overhead of actually processing a pixel and indexing into an image etc., the algorithm is implemented in about 10 lines of code. Its significance is:

- It is a theoretically sound solution to the throughput problem. Therefore, it questions the need for expensive hard to use parallel processing in high speed autonomy. It makes 20 mph autonomy feasible on throughput grounds and is four orders of magnitude more efficient than reactive approaches which process an entire 2 Hz image at that speed.
- It eliminates the image registration problem because the systematic errors which cause the problem vary little over the small width of the range window.
- It amounts to a computational solution to the stabilization problem within the limits of the sensor field of view. For limited field of view sensors, it provides an obvious basis for the generation of sensor pointing commands which keep the average range window centered in the image.
- Along with adaptive regard, it solves the sampling problem for practical purposes because the range ratio is very low over the small width of the range window. It theoretically guarantees throughput, response, and acuity within the limits of the vehicle and the sensor. It actively guarantees vehicle safety and adapts to vehicle speed, attitude and terrain shape implicitly.

3.3 Typical Example

In a typical image, the pixels that are actually processed form a jagged edged band across the horizontal. The width of the band decreases quickly as the vehicle speed increases and adaptive lookahead moves the window up the image. However, the validity of the **small incidence angle assumption**[4] guarantees that no matter what the terrain shape is and no matter what the vehicle attitude is, adaptive perception will generate a perfect wedge of geometry which is exactly the requirement for the current planning cycle.

The following figure gives a sequence of range images for a run of the RANGER simulator on very rough terrain using a simulated ERIM rangefinder where the pixels that were actually processed are highlighted as vertical white lines. On average, even in this worst case, only 200 range pixels out of the available 10,000 (or 2%) were processed per image. Thus, the 2% geometric efficiency of the sensor is effectively increased to 100% and throughput is increased by a factor of 50 times, or two orders of magnitude. The sparsity of the data in the planner window is interpolated away internally.

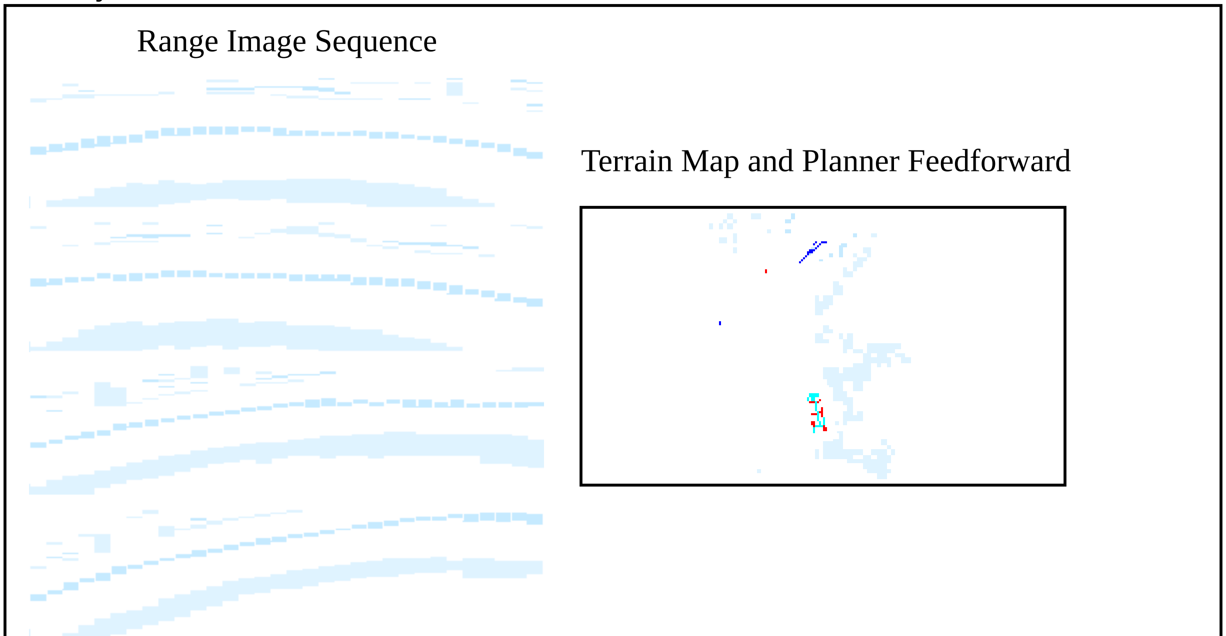


Figure 4 - Adaptive Rangefinder Perception

Notice that the algorithm adapts automatically to vehicle roll in the fourth image and that it processes the hill obstacle completely in the third image as soon as it appears. Notice also the skew of the vehicle icon with respect to the wedge of paths in the planner window. This is an aspect of the steering feedforward. The data is shifted up out of the image completely in the last image on the right side because of vehicle roll.

4. Of course, it is the relative accuracy of the small angle assumption across two images which really matters for throughput reasons, so the continuity assumption makes the small angle assumption even more valid.

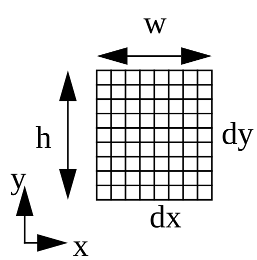
4. Terrain Map Management

Consider the implications of satisfaction of guaranteed response and acuity on the memory and computation required to manage a traditional terrain map. Up to 30 meters of lookahead is not uncommon and resolutions on the order of 1/6 meter are theoretically necessary. Using only 20 bytes of memory per map cell, over 1/2 megabyte of memory is required to store a typical map. If this map is stored as a physically coherent block of memory, it must be physically shifted and copied after the acquisition of each image. Using a realistic frame rate of 10 Hz, which is necessary to guarantee detection, the overhead involved in simply storing and managing the environmental model is not justified in a real time system.

4.1 Wrappable Map

The solution to this problem is a classical one from computer science - the ring buffer. However, there is no intrinsic requirement to insert and delete nodes into a linked structure because the map is spatially coherent and of uniform resolution. Hence, a simple array accessed with modulo arithmetic suffices to ***logically scroll*** the map as the vehicle moves.

The operating principle of the wrappable terrain map is that the indices into the array are determined by modulo arithmetic as follows:



$$map\left[\left\lfloor \frac{rem\left(\frac{x}{w}\right)}{dx} \right\rfloor\right]\left[\left\lfloor \frac{rem\left(\frac{y}{h}\right)}{dy} \right\rfloor\right] = z(x, y)$$

Figure 5 - Wrappable Map Indexing

The operator $\lfloor x \rfloor$ is the least integer function and $rem(x/y)$ is the floating point remainder function. These are implemented more efficiently than a function call in the code[5]. The operation of the technique when applied to three successive images is indicated below.

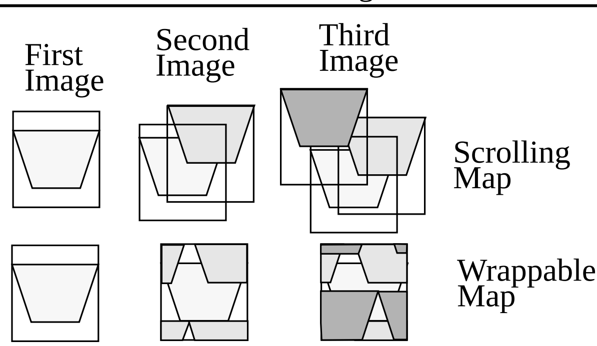


Figure 6 - Wrappable Terrain Map

This approach creates new problems, the most serious of which is that the mapping from world coordinates to map indices is multiply defined and therefore the inverse mapping is not a function. In mathematical terms, the coordinate transform is not **onto**.

5. By assuming excursions are limited to a few times the distance to the moon!

An infinity of points in global coordinates all map onto a single cell in the map, so remnants of images of arbitrary age may remain in the map indefinitely. Suppose the elevation at the point (15, 25) is needed and the map is 10 by 10. Then the point (5, 15) may also be in the map. A query for the elevation at (15, 25) may get the elevation at (5, 15) instead.

The system manages this problem in a very simple way. Although all data remains in the map until it is overwritten, each entry is tagged with the distance that the vehicle had travelled since the start of the mission at the time the pixel was measured. If the "age" of the last update is too old, the cell is considered to be empty[6]. This technique eliminates all of the overhead of map management except for the coordinate transformation necessary to access it. If pixels older than the length of the map are discarded in this way, it is impossible for old data to poke through the holes in new data. The impact is:

- it is never necessary to perform a copy of the map data structure
- graphical output had to be modified to wrap around in the map window

Theoretically, the complexity of map management is squared in the resolution unless the map is wrappable. The cost of management of this data structure is independent of both its size and its resolution.

4.2 Delayed Interpolation

The terrain map is not interpolated at all. This is done because the interpolation of the map requires a complete traversal and this is too expensive. Instead, the responsibility for interpolation is left with the users of the map. Spatial interpolation is wasteful because the ultimate use of the map is to evaluate vehicle safety and vehicle safety can be expressed as a time signal.

It is more efficient to ***delay interpolation*** until the point in the computation at which it is really needed, and this point occurs inside the tactical controller. Also, only a small portion of the map is actually used in some situations because the vehicle maneuverability is limited. Any interpolation of unused geometry amounts to a waste of resources.

A further aspect of the interpolation problem is that occlusion is inevitable anyway, so spatial interpolation can never succeed fully without unjustified and harmful smoothness assumptions on rough terrain.

Thus the tactical controller ***interpolates in time instead of in space.*** Throughout the internal processing of the tactical controller, the central data structure is a time signal which may or may not be known at a particular point in time. The system is robust by design to unknown signal values and, as a by-product of its processing, computes an assessment of how much geometry is actually unknown and reacts accordingly. In this way, interpolation and occlusion are treated in a unified way and the system considers too much of either to be hazardous.

4.3 Image Registration

A simple image registration algorithm is used in situations where the imaging density is reduced below the amount necessary to ensure that the geometry under the rear wheels comes from the same image as the front wheels in the feedforward simulation.

---

6. Age is measured by distance and not time because otherwise the geometry under the wheels eventually disappears when the vehicle stops.

This module recovers the vehicle excursion between images by matching the overlapping regions of consecutive images. Currently, only the elevation (z) coordinate is matched and this works best in practice due to systematic errors which are, as yet, unidentified. When the z deviation of two consecutive images is computed, it is applied to all incoming geometry samples in order to remove the mismatch error.

4.4 Fusion/Accumulation

After the mean mismatch error is removed, there are still random errors in the elevation data. The map manager computes mean and standard deviation statistics for map cells in situations when two or more range image pixels from the same image or sometimes from consecutive images fall into the same map cell. In order to do this, each map cell has an image frame number tag as well as the age tag. When the incoming frame number differs from the stored one, the statistical accumulators are zeroed.

The value of this mechanism in computing a statistically meaningful result is highly suspect. However, it survives because the deviations computed are a representation of the slope of the terrain within a single cell.

5. Adaptive Stereo Perception

Adaptive perception is also possible for stereo ranging systems. The basic principle of the **range window** can be converted to a **disparity window** for a stereo system because the two are related by the stereo baseline. There is a slight difference in the geometry of a stereo range image compared to a rangefinder image. The first is based on perspective geometry and the second is based on spherical polar geometry. Therefore, a disparity window corresponds to a window on the y coordinate and not the true polar range. However, in most circumstances, this distinction can be safely ignored.

5.1 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 maximum value. The value of the disparity at the maximum 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**.

5.2 Stereo Triangulation

The basic stereo triangulation formula for perfectly aligned cameras is quoted below for reference purposes. It can be derived simply from the principle of similar triangles.

$$\frac{X_L}{Y_L} = \frac{x_l}{f} \qquad \frac{X_R}{Y_R} = \frac{x_r}{f}$$

$$X_L - X_R = \frac{Y[x_l - x_r]}{f}$$

$$b = \frac{Yd}{f}$$

$$Y = \frac{bf}{d}$$

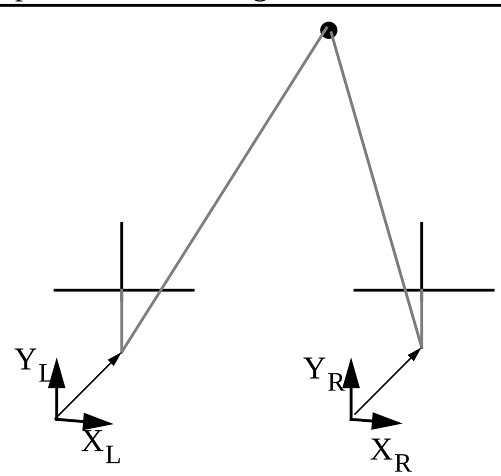


Figure 7 - Stereo Triangulation

5.3 Width of Disparity Window

Using the previous result, consider the width of the disparity window which corresponds to a typical range window. It is most useful to remove the dependence on the focal length by expressing disparity as an angle thus:

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

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

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

Thus, ***the range of disparities which correspond to a typical range window is very small indeed***. This implies that any process which robustly identifies global maxima of the disparity curve can generate the range window in an image. Again, because of the validity of the **small incidence angle assumption**, and because the window is based on the data itself, the process will automatically adapt for rough terrain and vehicle attitude. Such a process implements **adaptive sweep** inside the stereo algorithm.

5.4 Essential Difficulty

Consider the following typical correlation curve.

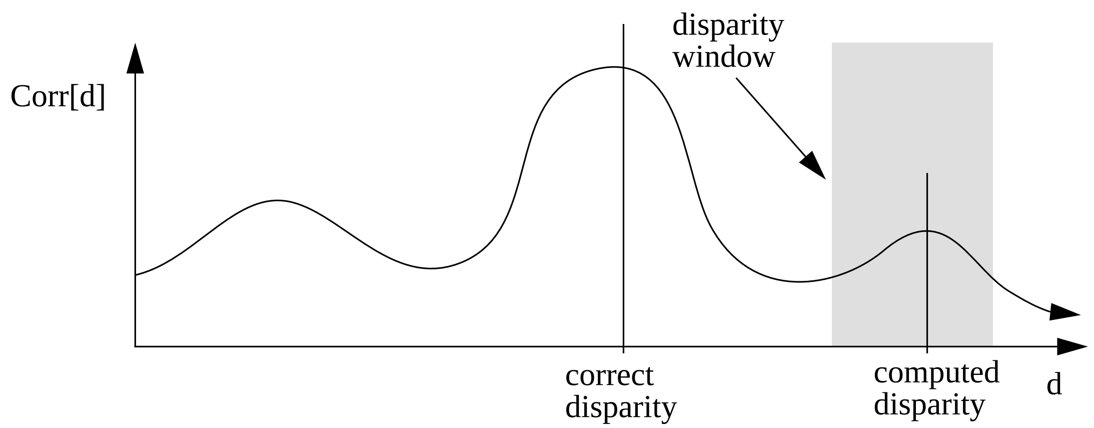


Figure 8 - Essential Difficulty

If the search for the maximum is confined only to the disparity window, then a **local maximum** will be found for the pixel, and not the correct global maximum. This implies that a typical image will contain pixels and regions where the ranging is incorrect as well as the correct range window as shown below:

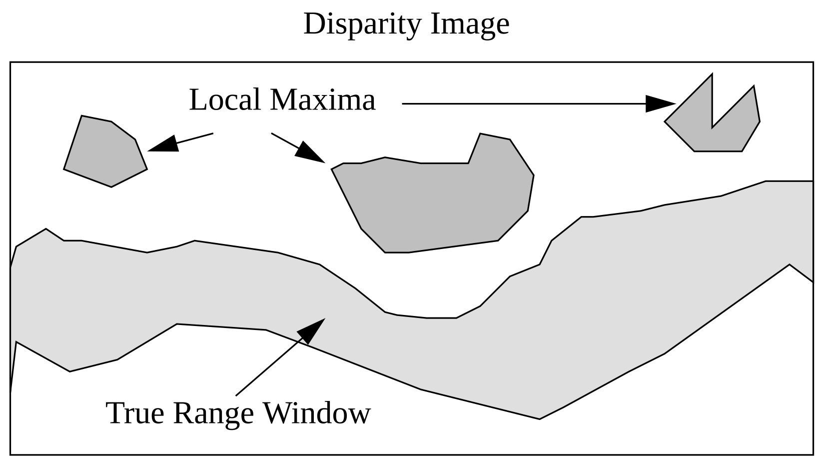


Figure 9 - Spurious Disparities

Spurious matches occur fundamentally because regions which do not correspond physically actually look more or less the same. This is called the **repetitive texture problem**.

5.5 Solutions to the Repetitive Texture Problem

A few techniques are available to alleviate the problem of repetitive texture in stereo.

5.5.1 Trinocular Stereo

A third camera or even more cameras help the situation because a spurious match is less likely to occur in all possible stereo pairs from a trinocular stereo head.

5.5.2 Left-Right Line of Sight

The correlation of left against right and right against left provides two separate correlation curves. A spurious match is less likely to show up in both curves.

5.5.3 Absolute Correlation

The use of normalized cross correlation as the matching criterion is superior to the use of the **sum of squared differences** or SSD criterion because the absolute value of the correlation coefficient actually has meaning. For example, a higher SSD may result from higher average brightness of pixels and may not imply a match at all. On the other hand, a local maximum in correlation curve can be assumed to be a global maximum when the correlation itself is high at the maximum.

5.5.4 Large Correlation Window

A large correlation window will suppress matches from small regions of repetitive texture. Unfortunately, this technique has the side effect of smoothing the range image as well.

5.5.5 Multidimensional Maxima

It can be argued that the peak of the correlation curve is only one of three possible partial derivatives. The partial derivatives of correlation across the image axes are also meaningful. For example, the partial across row index would be expected to be nearly flat as would the partial across the column index. Both assumptions fail at occluding edges, but occluding edges cannot be matched anyway so it is legitimate to discard them.

5.5.6 Monotone Range Assumption

The monotone range assumption can be used successfully in outdoor environments. It can be implemented by removing all matches in a column which are not part of the longest monotone run of disparity values.

5.6 Adaptive Scan

Adaptive scan is somewhat problematic in stereo because high angular resolution provides the texture necessary for accurate triangulation. Adaptive scan can be simply implemented in stereo by skipping columns in the input images. Note, however, that computation of correlation requires that the entire image be prefiltered. Adaptive scan can be implemented in the latter stages of stereo including, correlation, disparity and triangulation.

5.7 Typical Example

The following figures illustrate the operation of adaptive stereo on two input images. The initial input images appear at the top. The normalized, texture-enhanced images appear below the input images. The disparity image is shown to demonstrate the spurious matches which are a by-product of the disparity window approach. These correspond to local maxima in the correlation curve, but there is no information available to detect this. Finally, the cleaned up range image is presented. It incorporates an efficient filter based on the monotone range assumption which removes the local maxima and provides a clean range image for processing by the rest of the navigation system.

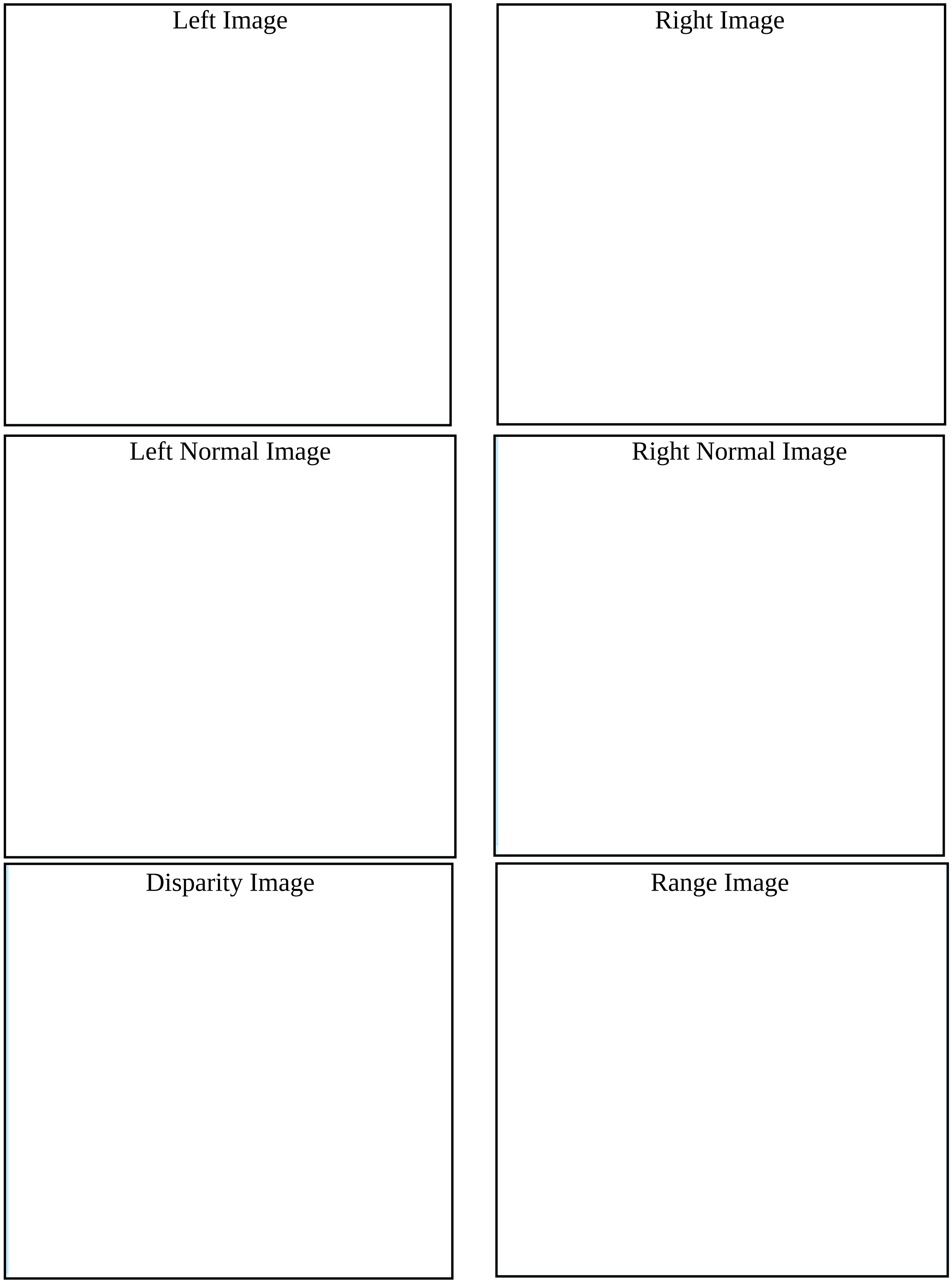


Figure 10 - Adaptive Stereo

6. Bibliography

[1] B. Brumitt, R. C. Coulter, A. Stentz, “Dynamic Trajectory Planning for a Cross-Country Navigator”, Proceedings of the SPIE Conference on Mobile Robots, 1992.

[2] T. S. Chang, K. Qui, and J. J. Nitao, “An Obstacle Avoidance Algorithm for an Autonomous Land Vehicle”, Proceedings of the 1986 SPIE Conference on Mobile Robots, pp. 117-123.

[3] M. Daily, “Autonomous Cross Country Navigation with the ALV”, Proceedings of the 1988 IEEE International Conference on Robotics and Automation, pp. 718-726.

[4] E. D. Dickmanns, “Dynamic Computer Vision for Mobile Robot Control”, Proceedings of the 19th International Symposium and Exposition on Robots, pp. 314-27.

[5] J. Gowdy, A . Stentz, and M. Hebert, “Hierarchical Terrain Representation for Off-Road Navigation”, In Proc SPIE Mobile Robots 1990.

[6] M. Hebert and E. Krotkov, “Imaging Laser Radars: How Good Are They”, IROS 91, November 91.

[7] M. Hebert. “Building and Navigating Maps of Road Scenes Using an Active Sensor”, In Proceedings IEEE conference on Robotics & Automation, 1989; pp.36-1142.

[8] 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

[9] B.K Horn and J. G. Harris, “Rigid Body Motion from Range Image Sequences”, Image Understanding, Vol 53, No 1, January 1991, pp 1-13

[10] R. Hoffman, E. Krotkov, “Terrain Mapping for Outdoor Robots: Robust Perception for Walking in the Grass”, Submitted to IEEE International Conference on Robotics and Automation, 1993.

[11] D. Keirsey,D. Payton, J. Rosenblatt, “Autonomous Navigation in Cross-Country Terrain”, proceedings of Image Understanding Workshop, 1988.

[12] A. Kelly, A. Stentz, M. Hebert, “Terrain Map Building for Fast Navigation on Rough Terrain”, Proceedings of the SPIE Conference on Mobile Robots, 1992.

[13] A. J. Kelly, “Essential Kinematics for Autonomous Vehicles”, CMU Robotics Institute Technical Report CMU-RI-TR-94-14.

[14] A. J. Kelly, “Modern Inertial and Satellite Navigation Systems”, CMU Robotics Institute Technical Report CMU-RI-TR-94-15.

[15] A. J. Kelly, “A Partial Analysis of the High Speed Autonomous Navigation Problem”, CMU Robotics Institute Technical Report CMU-RI-TR-94-16.

[16] A. J. Kelly, “A Feedforward Control Approach to the Local Navigation Problem for Autonomous Vehicles”, CMU Robotics Institute Technical Report CMU-RI-TR-94-17.

[17] A. J. Kelly, “A 3D State Space Formulation of a Navigation Kalman Filter for Autonomous Vehicles”, CMU Robotics Institute Technical Report CMU-RI-TR-94-19.

[18] A. J. Kelly, “An Intelligent Predictive Controller for Autonomous Vehicles”, CMU Robotics Institute Technical Report CMU-RI-TR-94-20.

[19] A. J. Kelly, “Concept Design of A Scanning Laser Rangefinder for Autonomous Vehicles”, CMU Robotics Institute Technical Report CMU-RI-TR-94-21.

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

[21] M. Marra, R. T. Dunlay, D. Mathis, “Terrain Classification Using Texture for the ALV”, Proceedings of SPIE Conference on Mobile Robots, 1988.

[22] L. Mathies, “Stereo Vision for Planetary Rovers”, Internation Journal of Computer Vision, 8:1, 71-91, 1992

[23] L. Mathies, S. A. Shafer, “Error Modelling in Stereo Navigation”, IEEE Journal of Robotics and Automation, Vol. RA-3, No. 3, June 1987.

[24] H. P. Moravec, “The Stanford Cart and the CMU Rover”, Proceedings of the IEEE, Vol. 71, Num 7, July 1983, pp. 872-884.

[25] K. Olin, and D. Tseng, “Autonomous Cross Country Navigation”, IEEE Expert, August 1991, pp. 16-30.

[26] M. Okutomi, T. Kanade, “A Locally Adaptive Window for Signal Matching”, in Proc ICCV, Dec 1990.

[27] M. Okutomi, T. Kanade, “A Multiple Baseline Stereo”, IEEE PAMI, Vol 15, No. 4, April 1993..

[28] B. Ross, “A Practical Stereo Vision System”, in proc CVPR 93.

[29] B. Wilcox et al. “A Vision System for a Mars Rover. SPIE Mobile Robots II”, November 1987, Cambridge Mass., pp. 172-179.

Index