TEC-0093

Active Stereo and Motion Vision for Vehicle Navigation

H. Keith Nishihara
Richard L. Marks
J. Brian Burns
Stanley J. Rosenschein

Teleos Research
2465 Latham Street, Suite 101
Mountain View, CA 94040

August 1998

U.S. Army Corps of Engineers
Topographic Engineering Center
7701 Telegraph Road
Alexandria, Virginia 22315-3864

US Army Corps
of Engineers
Topographic
Engineering Center

T

E

C

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
August 1998

3. REPORT TYPE AND DATES COVERED
Technical June 1994 - March 1996

4. TITLE AND SUBTITLE
Active Stereo and Motion Vision for Vehicle Navigation

5. FUNDING NUMBERS
DACA76-94-C-0018

6. AUTHOR(S)
H. Keith Nishihara Richard L. Marks J. Brian Burns
Stanley J. Rosenschein

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Teleos Research
2465 Latham Street, Suite 101
Mountain View, CA 94040

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Topographic Engineering Center
7701 Telegraph Road
Alexandria, VA 22315-3864

19. SPONSORING / MONITORING AGENCY REPORT NUMBER
TEC-0093

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This report describes work accomplished by Teleos Research in support of DARPA's Unmanned Ground Vehicle (UGV) Program. The presentation focuses on making forward-looking projections about opportunities for visual sensing on unmanned military vehicles based on the experience of this research program. A case is made for deploying UGV's following a small system model. Trends in commodity processor technology enable this possibility. The relationship between system cost and operational conservatism, system size and fragility, and speed and effort required, all support a shift towards smaller, cheaper implementations. We believe that this development model will rapidly take hold over the coming years. This report also presents a brief review of stereo performance characteristics relevant to the UGV mobility application. Several new techniques for enhancing stereo performance, including soft surface detection and disparity gradient compensation, are described.

14. SUBJECT TERMS
Unmanned Ground Vehicles, Stereo Vision, Real-Time Vision, Performance Trends, 3-D Texture

15. NUMBER OF PAGES
23

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT
UNCLASSIFIED

18. SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

19. SECURITY CLASSIFICATION OF ABSTRACT
UNCLASSIFIED

20. LIMITATION OF ABSTRACT
UNLIMITED

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

Contents

List of Figures iv

Preface v

1 Introduction 1
- 1.1 Alternative mobility sensing approaches 2
- 1.2 Stereo sensing characteristics 3

2 Stereo Cameras 5
- 2.1 FLIR Stereo 5
- 2.2 Low-Light (intensified) stereo 5

3 Stereo geometry 6

4 Ranging steeply inclined surfaces 8
- 4.1 Baseline length to height constraint 8
- 4.2 Skewed correlation window technique 9

5 3-D texture 11

6 Hardware/Software Performance Trends 14

7 Beyond Demo II 17
- 7.1 Small systems 17
- 7.2 Sensor agents 18
  - 7.2.1 Negative and Positive obstacles 18
  - 7.2.2 Wall follower 18
  - 7.2.3 Landmark tracker 18
- 7.3 Active vision 19
- 7.4 Structure from motion 19

8 Summary 21

List of Figures

1 Range resolution Wide FOV 6
2 Range resolution Narrow FOV 7
3 Field width versus range and focal length 7
4 Stereo imaging geometry 8
5 Skewed correlation window 9
6 Bush and rock image 13
7 3-D texture map of bush and rock image 13
8 Performance Trends 15

Preface

This research is sponsored by the Defense Advanced Research Projects Agency (DARPA) and monitered by the U.S. Army Topographic Engineering Center (TEC), under Contract DACA76-94-C-0018, titled, "Active Stereo and Motion Vision for Vehicle Navigation." The DARPA point of contact is Mr. Jon Bornstein and the TEC Contracting Officer's Representatives are Ms. Lauretta Williams and Mr. Thomas Hay.

ACTIVE STEREO AND MOTION VISION FOR VEHICLE NAVIGATION

FINAL REPORT
1 June 1994 to 31 March 1996

1 Introduction

This is the final report on work accomplished by Teleos Research on a three-year contract supported by DARPA's Unmanned Ground Vehicle (UGV) Program. This report focuses on the future of UGV sensing for mobility based on the experience of this funded research program.

The UGV Demo II program has deomonstrated that stereo sensing can be an effective tool for enabling UGV mobility. Sensor resolution, speed, and equipment costs still need to be improved over Demo II levels to achieve practical military deployment. This report intends to demonstrate that present day technology, if applied effectively, will support the deployment of UGV systems at practical cost levels. We believe, for reasons highlighted in this report, that the key to accomplishing this is to direct development efforts toward *small systems* that benifit from a number of multiplier effects once a minimum performance threshold is surpassed.

The following sections review the stereo sensing concept, its strengths and weaknesses, novel ways for applying stereo sensing to practical mobility tasks, and an analysis of performance trends in stereo systems. Section 7 presents a list of high-leverage development topics that would increase the effectiveness of stereo sensing on deployed UGV systems.

The UGV mission requires autonomously delivering a sensor package over land to designated locations, where terrain details must be sensed while the vehicle is enroute. Hazard avoidance and optimal path planning all depend on reliable feedback about the lay of the land. The UGV research program has investigated active and passive sensor technologies, and over its course it has produced a wealth of practical experience developing and applying those technologies. Among the sensing technologies investigated, binocular stereo sensing has yielded the best results at present and has been the primary focus of the Demo II research effort for supporting off-road mobility.

Under Demo II sponsorship, and in collaboration with the SRI and JPL stereo team members, Teleos carried out evaluations of stereo matching algorithms, and developed new algorithmic techniques for addressing the special needs of UGV mobility. Night stereo operation using passive FLIR and Intensified CCD cameras was successfully demonstrated. Teleos demonstrated operational feasibility through the development of real-time stereo sensing hardware prototypes. Portable implementations were assembled and characterized.

1.1 Alternative mobility sensing approaches

A mobility hazard detection mechanism must measure physical qualities that pose a navigation hazard directly, or indirectly, by finding environmental characteristics that are associated with the presense of an obsticle.

A major class of direct methods for detecting potential hazards are those that recover the 3-D scene geometry in front of the vehicle. These include active devices that emit radiation and measure time of flight, or angle of return, to estimate range to locations in the scene. These sensors can be scanned over a scene to assemble a range map which can then be interpreted to determine the navigability of the ground surface.

Active sensors have the major disadvantage of being a radiation emiter. For military systems, this exposes the user to detection and attack whenever they are employed.

Active range sensing techniques can be further classified according to the type of emission used. Sonar devices are useful proximity sensors, but do not have the spatial resolution necessary for identifying small ground shapes at driving lookahead distances. Multipath problems also limit performance.

RF and microwave imaging sensors have some nice properties, such as the ability to *see through* vegetation, such as tall grass. In addition to being emmiters, current systems have very low spatial resolution which limits them to close in proximity detection tasks.

Active light sensors using scanned lasers can produce high resolution range images for scenes at moderately large distances. In addition to being active, current systems are costly because of the precision optics involved and the need for high speed mechanical scanning of a high power laser.

Other approaches for using light emmisions to recover range have been proposed and are under development. One such system would eliminate the need for mechanical scanning by imaging short light pulses through a very fast electronic shutter that would allow sorting locations in the camera field of view according to their range. At present, however, this type of approach has shortcomings similar to those of the scanned laser systems.

Passive approaches to hazard detection have so far involved the use of imaging cameras of various types. Among these are multi-image triangulation techniques, the primary subject of this report. Generally, single- and multi-baseline stereo sensors have lower resolution at large distances, as compared with the best laser systems, however, they are passive and are generally less costly to build and operate. The next section gives a more detailed overview of the operating characteristics of stereo sensors.

Another passive image-based approach uses *scene analysis* techniques to classify regions in the image according to their material types using color and/or texture analysis. More sophisticated approaches would accomplish higher-level scene interpretation functions to identify/recognize physical objects, such as tree stumps, fences, buildings, or rock outcroppings. Such systems would be invaluable complements to range sensors. This is, however, a very difficult task, and in their current state of development, such high-level vision systems could not be used effectively, alone.

1.2 Stereo sensing characteristics

The binocular stereo imaging geometry uses triangulation to estimate range to imaged points that can be successfully matched across a pair of images. The following list reviews general properties of stereo sensors and several issues affecting their efficacy in the UGV mobility application:

1. Multiple images are required from separate locations to obtain stereo parallax. Section 7.4 discusses trends in stereo sensing research that will simplify stereo camera apparatus requirements.
2. Stereo range sensing is a passive process that can work with a variety of camera types including FLIR and intensified sensors. It will work with any imaging sensor that yields stable texture or edge markings on viewed surfaces. These texture markings are used to identify correspondences between stereo images that allow range computation by triangulation. Stereo performs best when there is abundant non-repeating texture present, as is the case in outdoor terrain. Performance is unreliable on targets with no texture, such as blue sky or reflected texture as reflections off of water. Section 2 describes several tests using night vision cameras.
3. Range accuracy is proportional to:

$$\frac{target\_distance^2}{baseline \times focal\_length \times sensor\_resolution}$$

   A consequence of this relationship is that stereo is most effective at close range, however, the square drop off of accuracy with distance can be compensated for by increasing lens focal length, camera baseline, and sensor resolution (number of pixels per field width). Section 3 gives numbers for typical sensor geometries.

4. Steep surface inclinations relative to the camera lines of sight are difficult to range using stereo correlation techniques. Typically, an inclination greater than about 30 degrees results in significant loss of correlation strength. Section 4.1 discusses a method developed under the Demo II program which extends the range of inclination that can be handelled by correlation-based stereo systems.

5. Stereo only detects visible surfaces. Tall grass and shrubbery that might be navigable appear as part of the detected topography and embedded hazards are not discernible through those covers. This limits the effectiveness of stereo sensing in regions with extensive ground foliage. Section 5 discusses an approach developed under the Demo II program for discriminating between diffuse and solid surfaces using stereo imagery.

6. Computational load is roughly proportional to number of measurements per second and number of range bins per measurement. These parameters are dictated by:

    (a) desired vehicle speed
    (b) stopping or maneuvering distance
    (c) path area that must be monitored
    (d) stereo sensor, interpretation, and vehicle control cycle time
    (e) minimum hazard size

    These elements are coupled. Generally, a higher vehicle speed increases the look-ahead distance necessary. Faster processing can reduce this distance, however there is a point of diminishing returns where vehicle dynamics, such as stopping distance, begin to dominate. Larger, look-ahead distances will generally involve monitoring a larger path area since the actual vehicle path is less certain. This leads to the general observation that there is a less than linear increase in processor load as sensor-control loop cycle frequency is increased. In some cases, processor load can acutally decrease with higher measurement cycle rates. Section 6 discusses trends in stereo processor performance that are relevant to UGV mobility.

2 Stereo Cameras

One of the key strengths of stereo sensing is that it is a passive process that can operate with a variety of imaging sensors. During daytime operation, a pair of low-cost CCD cameras are sufficient, and most research and testing of stereo has been with daylight cameras. Night vision stereo was thought to be feasible, however there was little experience using FLIR or intensified sensors prior to the Demo II program. Under the auspices of Demo II, night vision stereo configurations were tested and the results were positive.

2.1 FLIR Stereo

The availability of stable surface texture is a primary concern with using night vision sensors for stereo range finding. It was anticipated that FLIR images would show bland surfaces with low texture contrast. Tests carried out with a pair of Amber FLIR sensors on a stereo imaging mount showed significant texture contrast in outdoor scenes, including ground, grass shrubbery and trees. In a test during the first few hours after total darkness, texture contrast on materials such as a grass lawn were significantly higher than the texture contrast observed with a CCD camera on the same scene during daylight. Good stereo operation should be possible with FLIR as long as ground and air temperatures are not in perfect equilibrium. FLIR stereo was tested during daylight as well, with similarly good results regarding the texture contrast.

2.2 Low-Light (intensified) stereo

Stereo imagery from intensified CCD cameras were also tested under star light (and sky glow from distant city lights). The intensified cameras exhibited significant shot noise in the imagery. As would be expected, this noise increased dramatically as the scene illumination levels were reduced from lighting from nearby street lights, to lighting from distant street lights, and finally from star light. We found that stereo matching using large convolution and correlation operators could operate, to some extent, on these images in all but the lowest lighting level, however, the results were not as good as those obtained with FLIR stereo.

This result is a consequence of the relatively low surface texture contrast available in intensified images. We learned, in a related experiment, that the *sign correlation* algorithm, favored by Teleos for stereo processing, performs significantly better on these noisey images than all other stereo matching approaches evaluated in an extensive study carried out under the Demo II program.

3 Stereo geometry

Demo II stereo camera mounts have been tested with rigidly attached cameras, separated by baselines, varying from one meter down to 1/4 meter in length. These systems were pointed straight ahead of the vehicle and fairly wide angle lenses were employed to maintain visual coverage of the vehicle's path. The following plots (Figures 1 and 2) show optimal measurement resolutions for this camera configuration assuming that 1/3 pixel disparity resolution and 5 pixel image plane resolution can be be achieved by the stereo correlator. Observe that the range resolution degrades quadratically with distance while the transverse resolution degrades linearly. Note, also, that resolution improves linearly with lens focal length (and also with baseline length and CCD camera resolution).

Figure 3 shows the approximate area covered by lenses with three different focal lengths as a function of distance from the cameras. This highlights a fundamental tradeoff in stereo sensor head design. Larger field of view lenses reduce range acuity for a fixed camera resolution. As is discussed in section 7.3, this limitation can be dealt with by increasing sensor resolution or by employing multi-resolution cameras, or by allowing a head with larger focal length lenses to actively move.

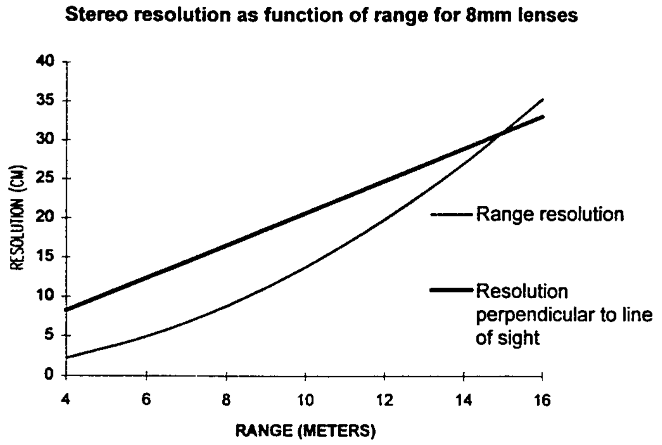

Figure 1: This figure plots estimated range and transverse resolution as a function of range to target using an 8mm wide field-of-view (FOV) camera lenses (approx. 90 degrees). A camera baseline separation of one meter is used and a 512 pixel scan line with 1/3 pixel disparity resolution is assumed. Transverse resolution is computed using the assumption that objects must be separated by more than 5 pixels.

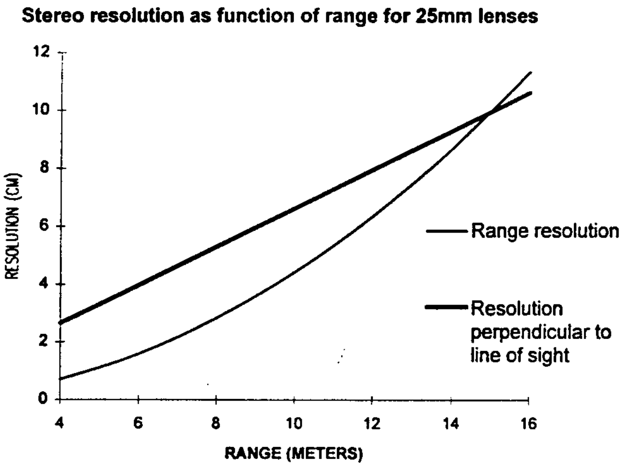

Figure 2: This figure plots estimated range and transverse resolution as a function of range to target using narrower 25mm lenses (approx. 35 degree FOV). As with Figure 1, a camera baseline separation of one meter is used and a 512 pixel scan line with 1/3 pixel disparity resolution is assumed.

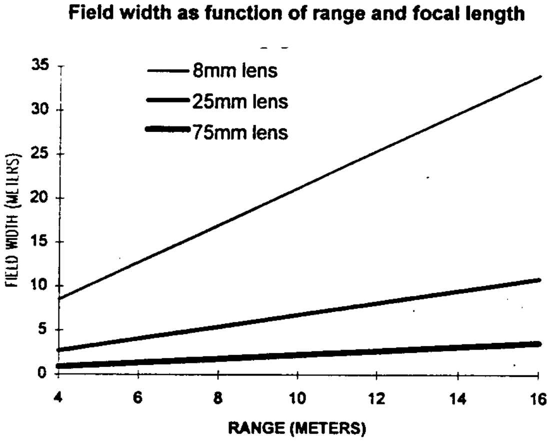

Figure 3: This figure plots estimated field width in meters as a function of range for three focal lengths.

4 Ranging steeply inclined surfaces

Surfaces that are inclined steeply relative to the cameras lines-of-sight give rise to large disparity gradients. These disparity gradients occur when the cameras view an inclined surface, such as the flat road out in front of the vehicle, as depicted in Figure 4. They can significantly affect the performance of area correlation based matchers because the receding surface under a correlation window does not register at any single disparity. This causes the correlation peak obtained to be lower and spread out, making detection of the peak more difficult and unstable.

4.1 Baseline length to height constraint

During the course of the Demo II program, we discovered a rather surprising result regarding stereo disparity gradients in the UGV imaging configuration, namely, that:

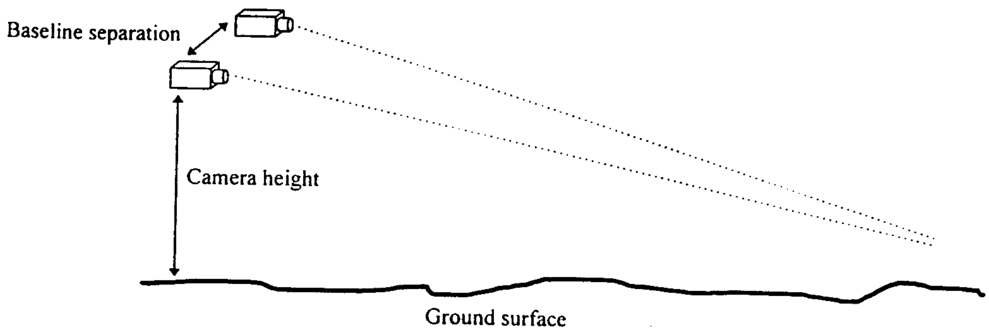

Figure 4: Stereo imaging geometry showing camera baseline separation and height of cameras above the ground surface. The ratio of baseline to height determines the stereo disparity gradient magnitude largely independent of other camera parameters, such as focal length.

$$\text{disparity gradient} \approx \frac{\text{baseline}}{\text{height}} \tag{1}$$

In other words, the disparity gradient depends primarily on the ratio of camera separation to camera height. It does not depend significantly on lens size, or pitch angle of the cameras, as long as the cameras are looking significantly farther ahead than they are high above the ground.

An important consequence of this result for the UGV program is the constraint it imposes on the baseline separation. Typical matching algorithms are seriously affected by gradients larger than about 0.2 pixels disparity change, per pixel in the image. This means that camera separation should not be larger than one fifth of the height of the camera head above the ground. Note that this constraint is in opposition with the desire to increase range acuity by increasing the camera baseline separation.

4.2 Skewed correlation window technique

In addition to designing the stereo imaging mount to limit the size of disparity gradients, there are ways to extend the performance of stereo matching algorithms to handle larger disparity gradients. One such technique developed under Demo II involves distorting the correlation windows to compensate for disparity gradients.

Compensation for vertical disparity gradients can be made when a correlation measurement is made by progressively shifting the horizontal disparity between the left and right correlation windows as we scan vertically over those windows, as shown in Figure 5. Adjusting the correlation window "skew" can greatly improve the correlation peak height obtained in images with large vertical disparity gradients. The window skew that gives the best correlation can also be used to directly estimate the local disparity gradient. A similar operation can be done to compensate for horizontal disparity gradients.

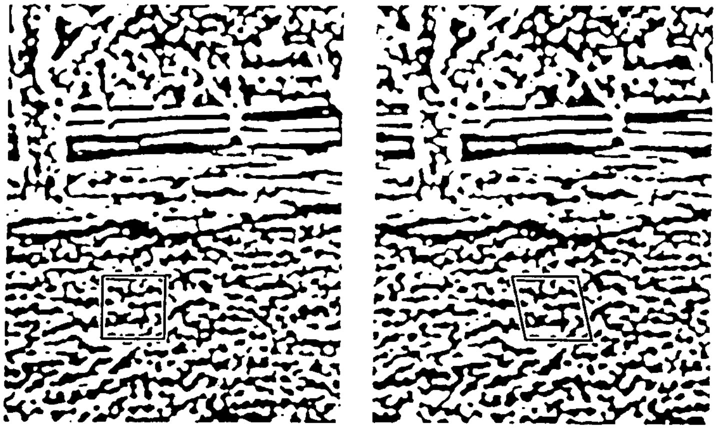

Figure 5: **A skewed correlation window is shown on the right that compensates for the vertical disparity gradient between the two images.**

The skewed correlation window technique has yielded almost a doubling of correlation peak height in tests on natural images with steeply inclined surfaces. This allows correlation based stereo matching to be applied on significantly steeper surface inclines. Algorithms employing skewed correlation windows to search over disparity, as well as over a range of disparity gradients, have been tested with good results.

5 3-D texture

Bushes and clumps of grass are difficult to distinguish from solid hazards in stereo range images. This reduces the effectiveness of passive stereo sensing in grass and shrub covered terrain, so any means for increasing the discriminability of *soft* objects would enhance UGV mobility. For example, if the path is blocked by range objects, but some section is determined to be grass or shrub, the UGV could be driven into that area in a slow *feel* mode to see if it can be traversed. Any hidden solid barrier or hole detected during the creep speed drive would cause the vehicle to take evasive action attempting to back out and move around.

Any ability to classify the makup of stereo range objects as being solid or soft would increase the effectiveness of the above scenario. One approach is to use color spectrum analysis to classify green vegetation from rock and dirt. This works during daylight, but not at night when FLIR stereo is employed. It also will not work as well when the vegetation color is not sufficiently distinct from solid hazards.

Under Demo II sponsorship, Teleos investigated methods for distinguishing solid surfaces from *soft* ones directly from the stereo correlation signal. It was observed that bushes and shrubs differ from solid obstacles like rocks in their diffuseness. This difference has an impact on the stereo correlation process which can be exploited to distinguish materials according to their solidness.

An example of this concept is illustrated in Figure 6. It shows a scene with a bush next to a rock. These objects look similar in many respects, including their color and size. They differ, however, in the way that they vary in depth. That is, their 3-D texture is different. The bush has large, high frequency variations in depth over its surface. The rock, on the other hand, varies more continuously in depth over similar spatial distances. This difference causes the quality of the stereo correlation to be much lower for the bush even when the correlation measurements are made at a coarse resolution.

Figure 7 shows a map of locations where high 3-D texture was located in the first image. Note that the silhouette of the bush shows up clearly.

This technique has several important properties:

1. No additional sensors required. The range texture approach to material classification does not require additional sensors since it operates off of the same imagery that is used by the stereo range finder.
2. No significant additional computation is required. The approach looks at the quality of the stereo correlation at different parts of the stereo image to determine material type.

This information is already being calculated as part of the stereo range calculation process.

3. It works with FLIR stereo for night operation. Since the classification is based on the stereo correlation, the technique can work at night using FLIR sensors. This is not the case for other approaches to material classification that rely on color or polarization effects.

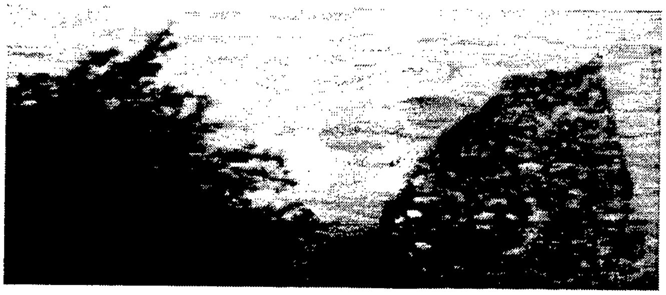

Figure 6: This image shows a scene with a rock and a bush. They look similar in many respects including their color and their range profiles. This makes it hard for a UGV range sensor to distinguish hard hazards from softer ones.

Figure 7: This is a map of locations in Figure 6 where the stereo correlation was abnormally low. This makes an effective classifier for bush-like objects. The low correlation is because of the bush's fine 3-D texture.

6 Hardware/Software Performance Trends

Up until the last year or two, special purpose hardware accelerators have been required to carry out the massive image processing computations required for recovering 3-D scene geometry using binocular stereo matching. For example, the Demo II hardware resources devoted to stereo included two high performance Datacube accelerator boards, a 68040 processor board, and a high performance SPARC computer.

Real-time vision processing speeds are limited by data movement bottlenecks and by available computation resources. Dedicated accelerators, such as Teleos' Prism-3 stereo and motion system[1, 2, 3], or JPL's DataCube stereo system[4], have been able to maintain high data throughput rates from digitizer to parallel processing pipelines, allowing intensive early vision computations to be carried out at video or near video rates.

However, the early advantage held by special board-level hardware over general purpose computers has been eroding over the past decade. Figure 8 illustrates the trends for a few examples of software and hardware stereo correlation systems. The figure compares representative systems by the number of correlations per second achieved by each. Among pure hardware systems, Prism-2[5] was an early instance that used conventional logic, such as adder and RAM chips, to implement a large kernel convolver and area correlator. Prism-3 used a similar architecture with more advanced Field Programmable Gate Array technology. This design ran at 4 times the clock rate and yielded almost an order of magnitude improvement in correlation speed. More recently CMU[6] developed a large piece of hardware that boosted performance by another two orders of magnitude. These data points suggest that hardware stereo correlators have been gaining about a factor of two in speed each year over the past decade.

Software stereo correlators on standard workstations, over the same period, have gone from being nearly four orders of magnitude slower [7] to just 1.5 orders of magnitude currently. If these trends persist for the remainder of the decade, software on personal computers will run essentially as fast as elaborate dedicated hardware systems by the turn of the century. There are several factors that give credibility to this somewhat paradoxical situation. First, and most significantly, clock speeds for board level accelerator designs have only risen by a factor of four or so in 10 years because of the physical limitations of clocking data onto pieces of wire. At the same time the instruction rate on microprocessors has risen by nearly three orders of magnitude during the same period. Closely coupled to this difference is the fact that investments made in making commodity processors faster, far outpace what can be justified for very low volume hardware accelerators for motion and stereo-correlation. Another important factor comes from improvements in processor bus bandwidth driven by the multimedia revolution. The current PCI bus on personal computers is specified to allow as many as ten live video signals to be moved simultaneously from video sources to host

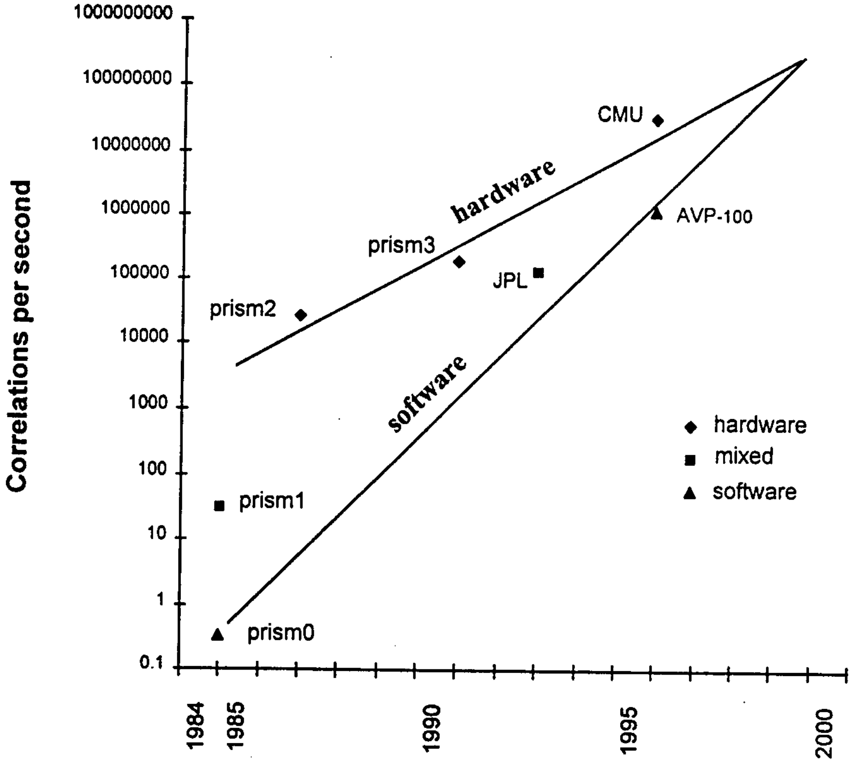

Figure 8: Software versus hardware performance trends for stereo correlation. Measurement rates for a representative set of stereo correlation systems developed over the past ten years are compared. The plot suggests that the performance gap between hardware- and software-based systems is closing steadily.

memory. This eliminates a critical bottleneck faced by earlier software implementations. Finally, most of the design techniques that have benefited hardware implementations, such as exploitation of separable convolutions, binomial approximations to the Gaussian, and boxcar filters, have helped also software implementations.

Based on these hardware versus software trends, Teleos Research has directed its technology development efforts toward PC-based software systems for early visual processing. The results have been favorable, and commercial sensing products are now on the market which carry out real-time tracking of people for teleconferencing, distance learning, and security applications. This tracking technology makes use of the same convolution and correlation algorithms employed under the Demo II program for stereo range finding.

As an example, a 66 MHz Pentium PC runs a motion-based figure detection algorithm at video rate, and drives an active camera head to follow subjects walking in an office environment. A 133 MHz dual Pentium system runs the same tracking software and computes stereo range images in parallel. As processor performance increases from year to year, this same software will be able to run at increasingly higher resolutions, and additional analysis modules will be able to run concurrently without affecting the low-level vision processing.

It is anticipated that multimedia enhancements expected in the next generation of Pentium processors will enable another quantum leap in visual measurement performance over the linear clock speed trend line.

7 Beyond Demo II

The UGV Demo II program fostered the development and testing of stereo sensing technology appropriate for vehicle guidance. As noted in the prior sections, stereo sensing offers a passive mechanism for detecting navigational hazards. This technology can be operated on mobile platforms fast enough to support real-time cross-country travel. Much has been learned about the operating envelope of the basic technique and methods developed to extend the performance envelope. Looking forward, a number of development areas present opportunities for solidifying the role of visual sensing technology in deployed UGV systems. We name these areas: (1) the small system view, (2) sensor agents, (3) active cameras, and (4) structure-from-motion.

7.1 Small systems

There are two opposed trends in autonomous mobile system development: the big system and the small system approaches. Until recently, because of processor limitations, the only viable option was to use larger systems. These computers and sensor pods were large and expensive, required a large power plant, air conditioning, and a large vehicle to carry them. As a consequence, mistakes could be very costly, leading to conservative operation and even greater investment in vehicle systems to enhance saftey.

As we move past the software versus hardware crossover point discussed in Section 6, the possibility of exploiting the opposite extreme of the big versus small system dimension arises. Once a small system can exceed a minimum performance level, a snowball effect the other way can occur. Cheaper systems do not have to do as much to be useful, and since they are more expendable, they do not have to operate as conservatively, making it easier to build them more cheaply.

As observed in the prior section, commodity laptop PC's now have sufficient power to carry out real-time stereo sensing; similarly, commodity frame grabbers, and cameras developed for the multimedia market are adequate for the UGV mobility task. These packages can be light and operate off of small batteries. This means the mobility platform can be small and capable of leveraging commodity technologies (e.g. golf carts) to deploy a practical UGV system.

This big system versus small system design choice has been explored extensively in the space program with unmanned Lunar and Mars rover designs. In the case of inter-planetary rovers, a key advantage of small expendable rovers is the opportunity for many more missions for the same dollar cost. This spreads the risk and even if the smaller rover systems have a higher failure rate, their greater number increases the likelihood of an overall mission success. The same rationale seems to apply to the UGV application.

7.2 Sensor agents

As visual sensor modules become cheaper, it becomes feasible to use more of them. One way to architect a system that does this is to employ a collection of measurement tools [8, 9] which are developed as specialist modules for accomplishing specific tasks in support of vehicle mobility. A short list of examples of such modules are:

7.2.1 Negative and Positive obstacles

Specialized modules can be developed to detect the different kinds of navigation hazards. In particular, negative obstacles, such as ditches or holes, present themselves differently from positive hazards, such as rocks and tree stumps. Dedicated sensing modules could be developed to attend to visual field locations, and process the stereo data to maximize the detection performance for each type of hazard.

7.2.2 Wall follower

Passive stereo can be used to guide a vehicle alongside a physical ground feature, such as the edge of a road cut or a tree or shrub line.[10] This would entail a side looking stereo sensor that monitors distance to the linear feature while the vehicle is in motion.

7.2.3 Landmark tracker

Large stereo range features at greater distances from the vehicle can often be detected sufficiently well to isolate them from the background.[11] These might be isolated trees, cliff sides, or mounds within sight of the vehicle. If they are sufficiently large, distant, and unique enough in appearance, they can be used as landmarks to monitor overground movement and vehicle position relative to those features. The precision of this technique could be enhanced using single-image pattern matching to reacquire an accurate directional fix on a previously seen target once the stereo system has localized its general position.

7.3 Active vision

A key limitation of the Demo II sensing head was its low spatial resolution. This was a consequence of the need for a wide field of view to fully cover the vehicle's path. The wide field lenses limited stereo range acuity in all three dimensions, as illustrated earlier in Section 3. There are several means for increasing range acuity while maintaining the necessary field of view. One is to use a higher resolution camera. This, however, is limited to about a factor of two improvement with available camera technology, and these non-commodity cameras are very expensive.

A second possibility is to use varying focal length lenses on the cameras. This could be accomplished with several camera pairs, each operating with a different field-of-view. The wide field-of-view system would operate as the present Demo II system did, detecting near-in hazards. Narrower field-of-view stereo modules would be directed farther ahead of the vehicle in the direction of its intended path. The narrow field-of-view sensor could detect problems earlier and allow time for evasive or corrective action. If this sensor were on an active mechanical mount, it would be possible to direct the sensing selectively to areas where the vehicle might be redirected. An active sensor head[12] would allow sensor and associated processing resources to be applied more efficiently.[13, 1] Provided that the sensor head is light, such a mount could react quickly and could be built economically.

7.4 Structure from motion

Teleos believes the "Holy Grail" of real-time stereo sensing for mobility applications will be the development of single-sensor structure-from-motion technology. This will ultimately extend the operating range of stereo systems to very large distances, further reducing the cost of operation of such systems.

Current stereo sensing operates with a set of two or more images taken from cameras on a carefully calibrated mounting frame that fixes the baseline directions and lengths. In motion-based stereo, a single camera in motion can be used to recover range structure over time. This temporal dimension is a source of information that has not been tapped in current designs because of the high processing demands of real-time motion algorithms.

Structure-from-motion using sequential images from a single camera is similar to stereo range computation, but with the added problem of computing the motion of the camera between images. Knowing this camera motion is critical to interpreting absolute range. Generally, this problem can be solved up to a scale factor provided sufficiently many point correspondences are known between sufficiently many successive images. The numbers required are not very large (minimum of five or so points in two images), however, the sensitivity to error can be large when the numbers are small. As the numbers of points and images increase, precision will rise, however, the computational demands also rise rapidly.

As the balance of processor power shifts, we expect to see structure-from-motion measurement supplant static stereo technology in many application areas. The steady rise in available processor resources will give rise to another snowball effect that relates to motion computation. This is the measurement rate versus search range relationship. It takes a linear increase in processor power to increase the measurement rate. A linear increase in measurement rate gives rise to a linear reduction in motion search range required to maintain track of an image feature in an image moving because of vehicle motion. This linear reduction in the 2-D search area gives rise to a squared reduction in processor resources required. Thus, once a minimum processor threshold is surpassed, fast operation can be cheaper than slow operation, therefore rates ultimately will be limited by camera frame rates.

Single-camera structure-from-motion computations have one disadvantage: the structure is computed up to a scale factor. This missing scale factor can be recovered in various ways, e.g., if the vehicle movement over land is measured acurately, this can be factored into the analysis as baseline information. Alternatively, calibration can be acquired from known landmarks. For example, as a vehicle drives forward, the approximate distance to the ground surface, just being occluded by the hood, can be estimated and used to set the scale for the range surface being computed.

This approach fosters a number of effects that enhance the range relief of potential hazards, such as ditches and holes, that are particularly difficult to detect in stereo imagery. For example, a forward looking sensor will see a build-up of detail over time as the vehicle drives over the scene it is monitoring. An extended temporal analysis of this continuous imagery will yield much higher range acuity. The vehicle driving path also can be adapted to enhance the performance of a structure-from-motion analysis. For example, a side to side weaving course would create motion shears at image locations corresponding to range discontinuities making them detectable at larger distances.

Structure-from-motion sensing paired with binocular stereo to obtain absolute range calibration could provide the best of both techniques in an efficient passive-sensor package.

8 Summary

A case has been made in this report for deploying UGV's following a *small system* model. Trends in commodity processor technology enable this possibility. The relationships between system cost and operational conservatism, system size and fragility, and speed and effort required, all support a shift towards smaller, cheaper implementations. We believe that this development model will rapidly take hold over the coming years. This report also presented a brief review of stereo performance characteristics relevant to the UGV mobility application. Several new techniques for enhancing stereo performance, including *soft* surface detection and disparity gradient compensation, were described.

References

[1] H. K. Nishihara, H. Thomas, and E. Huber. Real-time tracking of people using stereo and motion. In *Machine Vision Applications in Industrial Inspection II, Benjamin M. Dawson, Stephen S. Wilson, Frederick Y. Wu Editors*, volume Proc. SPIE 2183, pages 266–273, 1994.

[2] H. K. Nishihara and S. J. Rosenschein. Stereo machine vision in fpga's. In *PLD Con95*, pages 25–27, Santa Clara, CA, April 1995. CMP Publications Inc, Manhasset NY.

[3] R. Marks, S. Rock, and M. Lee. Real-time video mosaicking of the ocean floor. In *Proceedings of IEEE Symposium on Autonomous Underwater Vehicle Technology*, Cambridge, MA, July 1994.

[4] 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.

[5] H. K. Nishihara. Real-time implementation of a sign-correlation algorithm for image-matching. Technical Report TR-90-2, Teleos Research, 1990.

[6] S. Kimura, H. K. T. Kanade, A. Yoshida, E. Kawamura, and K. Oda. Video-rate stereo machine. In *Mobile Mapping Symposium*, Columbus, OH, May 1995.

[7] H. K. Nishihara. Practical real-time imaging stereo matcher. *Optical Engineering*, 23(5):536–545, October 1984. Also in Readings in Computer Vision: Issues, Problems, Principles, and Paradigms, edited by M.A. Fischler and O. Firschein, Morgan Kaufmann, Los Altos, 1987.

[8] S. Ullman. Visual routines. Technical Report AI Memo 723, Massachusetts Institute of Technology, Cambridge, MA, June 1983.

[9] H. K. Nishihara. Minimal meaningful measurement tools. Technical Report TR-91-01, Teleos Research, 1991.

[10] R. Marks, H. Wang, M. Lee, and S. Rock. Automatic visual station keeping of an underwater robot. In *Proceedings of IEEE Oceans 94 Osates*, Brest, France, September 1994.

[11] J. B. Burns and H. K. Nishihara. Feasibility of using stereo to assist navigation. Technical Report TR-92-02, Teleos Research, Mountain View, CA, 1992.

[12] M. J. Swain and M. Stricker. Promising directions in active vision. Technical Report CS 91-27, University of Chicago, Chicago, IL, 1991.

[13] H. K. Nishihara, H. Thomas, E. Huber, and C. A. Reid. Real-time tracking using stereo and motion: Visual perception for space robotics. In *Proceedings of the Third International Symposium on Artificial Intelligence, Robotics, and Automation for Space*, Pacadena, CA, 1994.