ARMY RESEARCH LABORATORY

Toward High Resolution, Ladar-Quality 3-D World Models Using Ladar-Stereo Data Integration and Fusion

by William F. Oberle and Lawrence Davis

ARL-TR-3407 **February 2005**

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

DESTRUCTION NOTICE—Destroy this report when it is no longer needed. Do not return it to the originator.

Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5066

ARL-TR-3407 **February 2005**

Toward High Resolution, Ladar-Quality 3-D World Models Using Ladar-Stereo Data Integration and Fusion

William F. Oberle
Weapons and Materials Research Directorate, ARL

Lawrence Davis
University of Maryland

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED (From - To)
February 2005	Final	December 2003 to August 2004

4. TITLE AND SUBTITLE

Toward High Resolution, Ladar-Quality 3-D World Models Using Ladar-Stereo Data Integration and Fusion

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

William F. Oberle (ARL) and Lawrence Davis (Univ of Maryland)

5d. PROJECT NUMBER
622618AH03

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

U.S. Army Research Laboratory
Weapons and Materials Research Directorate
Aberdeen Proving Ground, MD 21005-5066

8. PERFORMING ORGANIZATION REPORT NUMBER

ARL-TR-3407

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

An approach and architecture to incorporate data integration and fusion of laser radar (ladar) and stereo data to generate high resolution, ladar-quality three-dimensional world models is described. Our primary interest involves complex environments that have proved difficult for traditional stereo algorithms to produce accurate information. The principal novelty of our work is the use of ladar information as *a priori* disparity information. Initial results verify the validity of the approach. Improvement in the identification of occluded regions and reduction of the error in disparities are observed.

15. SUBJECT TERMS

data fusion; data integration; disparity; ladar; stereo vision; world model

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			William F. Oberle
Unclassified	Unclassified	Unclassified	UL	35	19b. TELEPHONE NUMBER (*Include area code*) 410-278-4362

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Contents

List of Figures

List of Tables

Acknowledgments

The authors would like to thank Thomas Haug of the U.S. Army Research Laboratory (ARL) for his time and effort in reviewing the report. His comments and suggestions proved valuable in the preparation of the final version of the report.

This work was the outgrowth of research performed under Cooperative Research and Development No. 0312-A-C799 between the University of Maryland and ARL.

INTENTIONALLY LEFT BLANK

1. Introduction

A robotic system depends on a variety of on-board sensors providing information concerning its environment in order to accomplish required mission objectives. Examples of typical mission objectives for robotic systems are autonomous mobility and object detection. In general, most of these sensors are capable of very accurately perceiving only a narrow aspect of the environment. For example, ladar[1], radar, and sonar sensors provide depth and displacement information, while infrared (IR) sensors provide data about thermal emissions within the environment. On the other hand, machine vision systems using a "daylight" camera(s) can be one of the most informative sensors, providing information across a wide range of sensor modalities (e.g., color, shading, texture, etc.). As Bischoff and Graefe (1998) observed, "Vision is the most powerful sensor modality for providing rich and timely information on a robot's environment." Unfortunately, the versatility of the vision system information is often accompanied by the complexity of the data analysis. Even a seemingly simple question such as the color of an observed object is confounded by factors such as illumination (i.e., the color consistency problem), which often lead to inconsistent results.

Faced with this disparity in the information provided by the sensors about robotic systems, we naturally exploit the possible benefits offered by the "integration" or "fusion" of data from multiple sensors to construct a broader and more inclusive model of the robot's environment. While multi-sensor data fusion appears to be a common approach in the target recognition and automatic target recognition communities (e.g., Hall, 1992; Stevens, Beveridge, and Goss, 1997), much less work is reported in the literature about data fusion relative to constructing the environment of a robotic system. The work by Abidi and Gonzalez (1992) provides an introduction to the subject. Of interest to us in this research is the fusion and integration of ladar sensor data and imagery data from a stereo camera pair.

Output of a ladar sensor is range (distance) information based on the time of flight of a laser pulse emitted by the sensor that is reflected off an object and back to the sensor. Thus, the range information is a direct measure and is generally accurate. Figure 1 provides typical ladar range data for real-time ladar sensors represented as an image using false color to quantify range. As can be observed in the figure, the resolution of the range image is substantially less than that available with most camera data. In addition, the ladar data suffer from what is known as the "mixed point problem" (Dias, Sequeira, Goncalves, & Vaz, 2001). Essentially, the mixed point problem results from the fact that the laser pulse has a non-zero width (i.e., appears more like a disc than a point). At edges or depth discontinuities in the scene, the laser pulse reflects from objects in both the foreground and background. In this case, the measured distance is a

[1] An acronym of laser detection and ranging, ladar uses laser light for detection of speed, altitude, direction and range; it is often called laser radar. See the photonics dictionary – web site: http://www.photonics.com/dictionary/.

combination of the distances to foreground and background objects. As a result, edges often tend to exhibit sawtooth-like patterns, several instances of which are evident in figure 1. The image of the same scene as viewed from the left camera of a stereo camera pair is shown in figure 2. Clearly visible are many features not present in the ladar range image (e.g., shadows, color, and clear boundaries at depth discontinuities). Camera data are a measurement of the energy (intensity) of reflected light off object surfaces and changes depend on the scene illumination. Thus, most of the information obtained from camera data is the result of some form of analysis. The most important derived information for a stereo camera pair is the three-dimensional (3-D) reconstruction of the scene (i.e., world model) via a geometric analysis. For relatively "simple" environments[2], both ladar and stereopsis tend to provide acceptable results. However, the same is not necessarily true for more "complex environments".

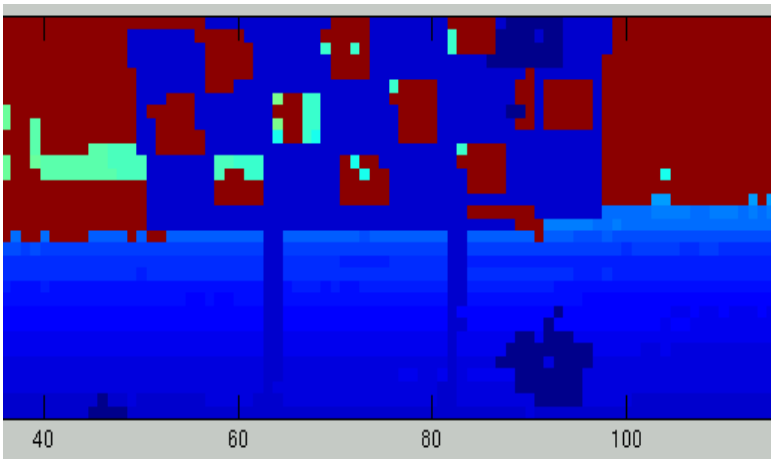

Figure 1. False color ladar range image (Oberle & Haas, 2002).

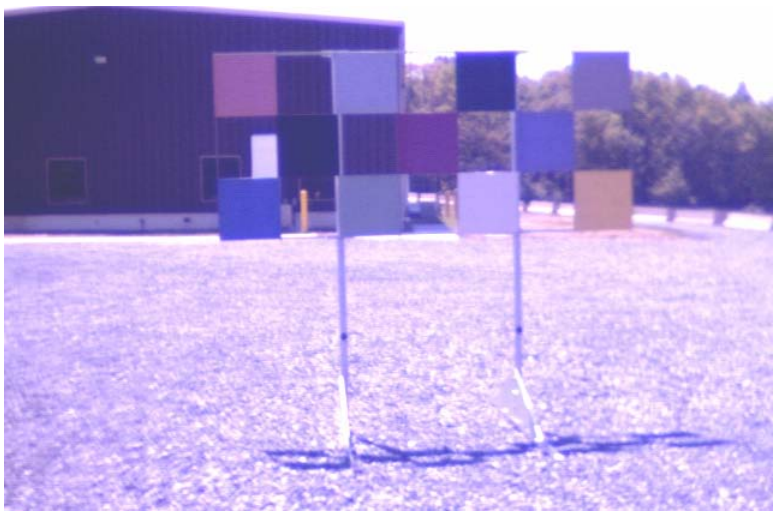

Figure 2. Left-hand camera image from stereo camera pair of ladar scene in figure 1 (Oberle & Haas, 2002).

[2]A "simple" environment is one in which there are few depth discontinuities and object surfaces tend to be smooth.

To summarize, ladar data consist of accurate spatially low resolution range data while stereo camera data consist of spatially high resolution but generally noisy reflectance data from which range data can be derived. In addition, the stereo camera data provide scene information not available with the ladar (e.g., color or sharp depth discontinuities). The research described in this report focuses on improving the calculated world model of complex environments through the use of data integration and fusion (Abidi & Gonzalez, 1992) of ladar sensor data and stereo camera imagery. Our specific research objectives are to

1. Improve the solution to the stereo correspondence problem[3] and by extension, the 3-D stereo reconstruction[4] problem by using data integration[5] with ladar range data as *a priori* disparity information; and
2. Improve the 3-D world model through the data fusion[6] of the improved 3-D stereo reconstruction information with ladar data.

A number of researchers have used data fusion of ladar and vision data to enhance the 3-D world model. For example, Dias, Sequeira, Goncalves, and Vaz (2001) used fusion to address the mixed point problem for both indoor and outdoor environments. Nickels, Castano, and Cianci (2003) presented a unified architecture for fusing lidar[7] and stereo range data to create a summary map of obstacles and free space surrounding a robot. Spero and Jarvis (2002) detailed their efforts to fuse imagery data and ladar to construct a high resolution model of the environment in terms of surface shape and color (a common approach to obstacle detection and tracking in the unmanned ground vehicle community (e.g., see Chang, Hong, Rasmussen, & Shneier, 2002). However, our literature review yielded no research addressing the use of the ladar data as *a priori* information to improve the solution of the stereo correspondence problem.

The purpose of this report is to describe a proposed approach to accomplish the research objectives as enumerated and to detail the current status of the research. In section 2, a proposed architecture to accomplish the stated objectives is presented and discussed. Section 3 describes relevant characteristics of the application domain (i.e., complex environment) and their influence on the selection of the "stereo matching" algorithm (used to solve the stereo correspondence problem) selected for the initial proof-of-concept experiments. Results of these experiments are provided in section 4. Finally, in section 5, a summary and outline of future work are presented.

---

[3]Correspondence Problem: "Which parts of the left and right images are projections of the same scene element?" (Trucco & Verri, 1998)

[4]Reconstruction Problem: "Given a number of corresponding parts of the left and right images, and possibly information on the geometry of the stereo system, what can we say about the 3-D location and structure of the observed object?" (Trucco & Verri, 1998)

[5]Synergistic use of sensor data to accomplish specific task.

[6]Combining data to generate a single model representation.

[7]Another acronym of laser detection and ranging, same principle as ladar.

2. Architecture

A flow diagram of the necessary steps to accomplish the research objectives is shown in figure 3. As illustrated in the figure, the processes are separated into five different functional layers. Within the pre-processing and data integration layers, individual algorithms necessary to achieve the research objectives are identified. At the present time, a number of the algorithms in these layers have been completed. Much less specific is the data fusion layer; algorithms in this layer will augment the work of Dias et al. (2001); Nickels, Castano, and Cianci (2003); Spero and Jarvis (2002); Chang et al. (2002), and others as we continue our work. A discussion of the first three functional layers follows.

2.1 Input Layer

The ladar and camera data listed as input in the input layer of figure 3 represent the minimal input required to perform the analysis. Additional input that would be useful to the analysis include range or intensity images from the ladar and left-right camera registration. Sources of input error that can propagate throughout the analysis involve the camera calibration information and the 3-D ladar data. On a moving vehicle, especially over rough terrain, vibrations can result in changing camera settings that affect the camera calibration. Although range is a direct measurement of the ladar sensor, 3-D coordinate data are a derived measure. Essentially, the ladar system uses a spherical coordinate system (ρ,θ,ϕ) in determining the 3-D coordinates. The range, ρ, is directly measured while the two spherical coordinate angles, θ and ϕ, are associated with the location of the laser emitter. For real-time systems, the angles and the emitter location are often based on a pre-operation calibration.[8] This calibration, as with the camera calibrations, could change during periods of operation.

2.2 Pre-Processing Layer

The pre-processing layer is optional if the ladar sensor and cameras are registered off line and the registration does not change during operation, i.e., the ladar sensor and cameras are rigidly mounted and move as a single unit. If this is not the case, then three registrations (ladar-left camera, ladar-right camera, and left-right camera) are required in order to complete the analysis. However, given any two of the registrations, the third can be determined. The pre-processing layer as shown assumes that the cameras are not registered. If the cameras are registered, then only the ladar-left or ladar-right camera registration needs to be determined. In either case, it is necessary to determine a set(s) of matching, corresponding, or homologous points between the 3-D ladar data and the 2-D image data (left, right, or both cameras). If the input from the ladar sensor includes either a range or intensity image, the image can be used in the matching.

[8]Private communications, G. Haas, U.S. Army Research Laboratory, April 2004.

Otherwise, a range image for the ladar data must be created. A number of procedures can be used to create the range image, the simplest being to use false color to quantify the range and to use the scanning properties of the ladar (e.g., number of emission per horizontal line, number of vertical positions, etc.) to define the image size.[8] Figure 1 is an example of this approach.

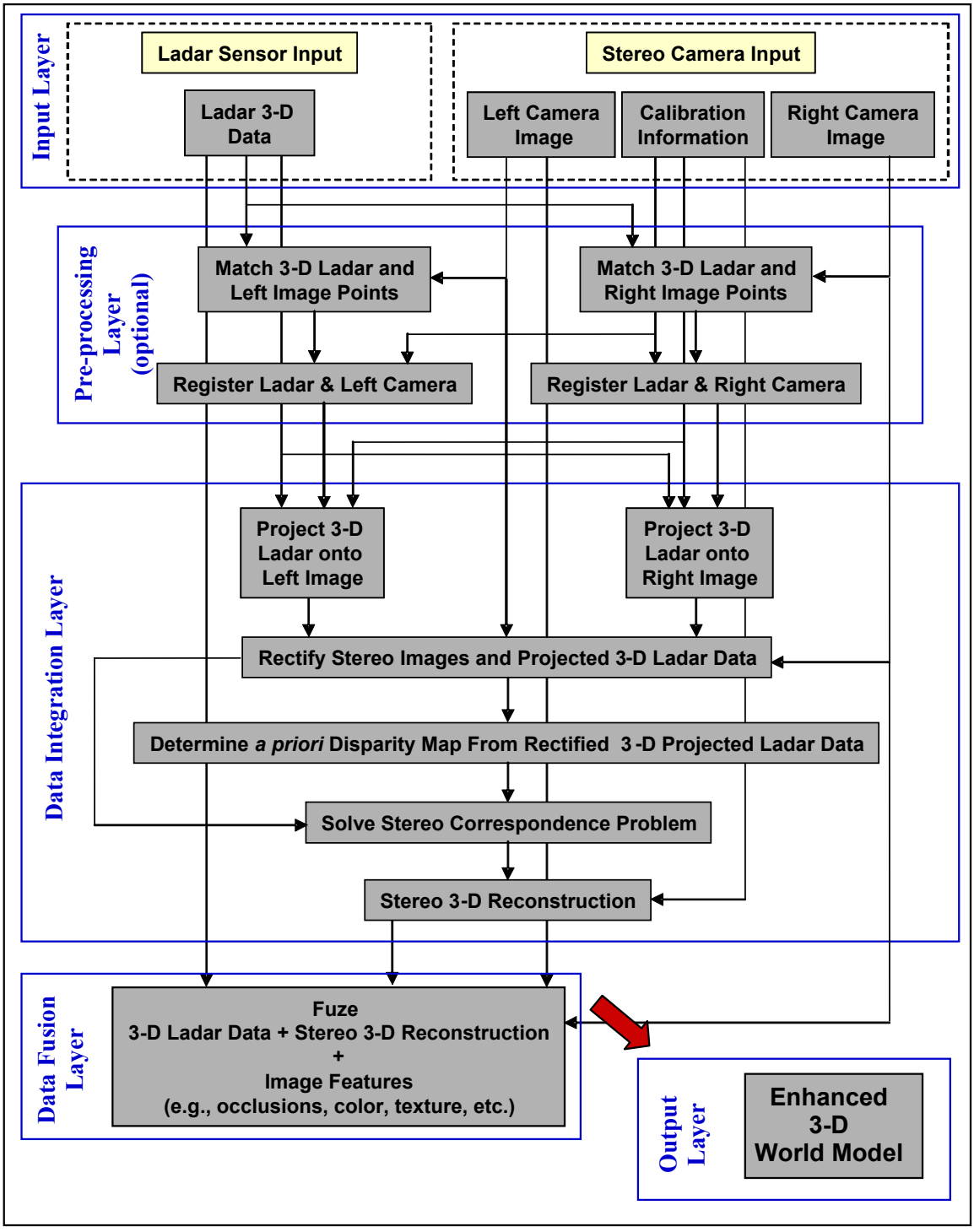

Figure 3. Proposed architecture for research effort.

Once a ladar image (intensity or range) is available, details of the matching algorithm must be addressed. The most direct approach is to manually select the corresponding points. Although potentially time consuming, this approach should result in a relatively accurate set(s) of matching points and is the approach used in this work. However, this approach is only applicable if the registration(s) need to be calculated infrequently. If the ladar and cameras are not rigidly mounted, the registration(s) will have to be performed on a continuous basis and the matching process will have to be automated. Corners (Elstrom, 1998) or edges (Dias et al., 2001) are two common features frequently used in the determination of corresponding points. However, since one of the images is based on ladar range or intensity and the other is based on a camera, care must be taken in using an automated evaluation of the "goodness" of the match. Standard correlation techniques may not be viable in these circumstances.

Ladar-camera registration involves determining the rigid body transformation between the ladar and camera coordinate systems based on a matched set of 3-D (ladar) and 2-D (camera) coordinates. A comparison of approaches to accomplish this calculation has been performed with several acceptable methods identified (Oberle & Haas, 2004). Two methods, one by DeMenthon and Davis (1995) and another by Bouguet (2003) were used.

2.3 Data Integration Layer

It is within the data integration layer that the major effort of our work has been focused to date. The principal novelty of this work is the integration of 3-D ladar information as an *a priori* image disparity map to improve the solution to the stereo correspondence problem.

The first step in creating the *a priori* disparity map from the 3-D ladar data is to project the ladar data onto the left and right camera images while simultaneously building a table of left and right image pixel pairs that are images of the same 3-D ladar point. A pin-hole, projective camera model is used to perform the projections. As each 3-D ladar point is projected onto the left and right camera images, a "partial mapping" between left-image pixels and right-image pixels is generated. The mapping is termed partial since not every pixel in either the left or right image is guaranteed to be in the range of the projection of the 3-D ladar points. Although the calculation of the projection of a 3-D point onto an image is straightforward, the overall mapping must be constructed in such a way to ensure that the resulting correspondence between pixels in the left and right images is unique. Given two 3-D ladar points, there are four possible results for the mapping as illustrated in figure 4. In cases 1 and 4, the correspondence between the image pixels is unique. However, in cases 2 and 3, a single pixel in the left (right) image corresponds to two pixels in the right (left) image. To resolve this ambiguity, the 3-D ladar point with the greatest range and the associated pixel correspondence is discarded. The motivation for this decision is based on the fact that if two 3-D points map to the same point in an image plane, only the closest point is visible with the other point being occluded. Code to perform this particular algorithm has been completed.

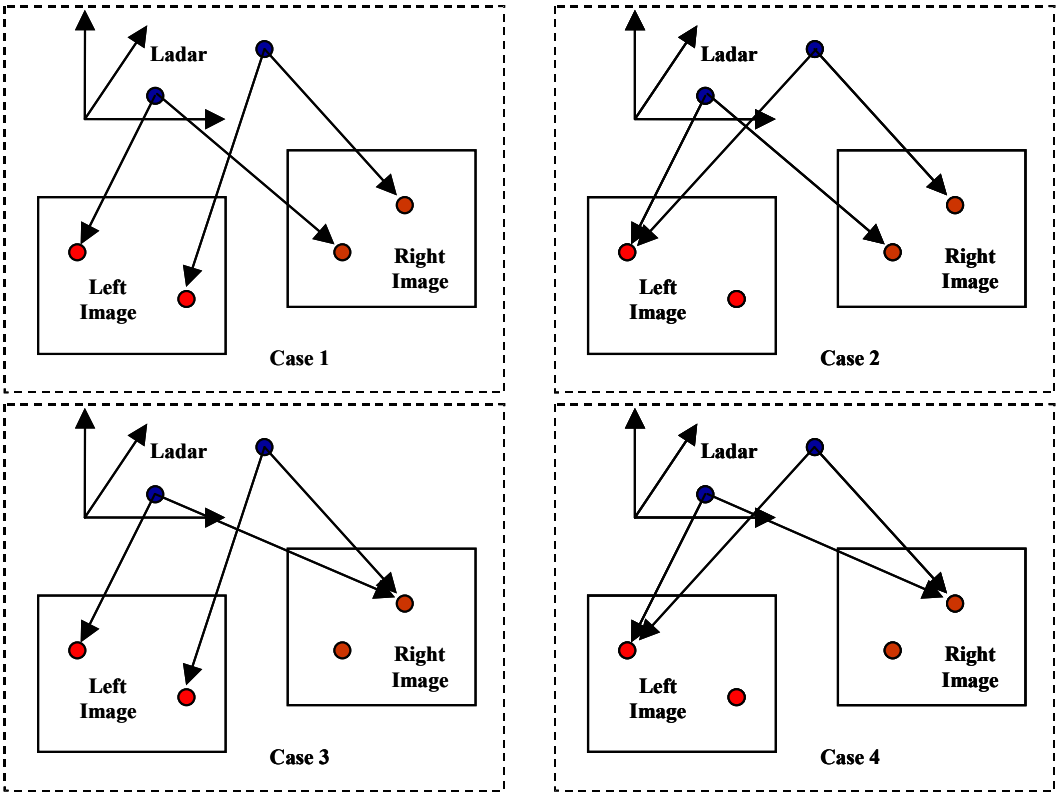

Figure 4. Possible outcomes of mapping two 3-D ladar points onto left and right camera images.

In order to reduce computation time, algorithms designed to solve the stereo correspondence problem expect the stereo image data to be rectified. This is also true for all the algorithms undergoing consideration for use in this work. Thus, the next step in computing the *a priori* disparity map and preparing to solve the correspondence problem is rectification of the image data and adjustment of the associated pixel correspondences determined in the first step just discussed. Essentially, rectified images are ones in which the same scene elements are on the same horizontal scan line in each image. If the cameras of the stereo camera pair are calibrated and registered, then rectification can be accomplished by the mapping of the left and right images to new images whose coordinate systems are related by only a translation along the x-axis. This approach is described in almost every book on computer vision (e.g., Faugeras, 1993, or Trucco & Verri, 1998). Unfortunately, if the camera calibration or registration information is erroneous, the rectified images tend to be vertically shifted (Oberle, 2004). As mentioned earlier, because of vehicle vibration, the camera calibration and/or registration information will most likely change during operations. Thus, for this work, a rectification algorithm not dependent on either camera calibration or registration information will be implemented. The algorithm will be based on the work of Lim, Mittal, and Davis (2004); Pollefeys, Koch, and Van Gool (1999); and Loop and Zhang (1999). At the same time that the images are rectified, the table of left-right pixel correspondences is adjusted to remain consistent with the rectified images.

Once the images are rectified and the table of left-right pixel correspondences is adjusted, the *a priori* disparity map can be constructed. If $p_1 = (x_1, y_1)$ represents the coordinates (in pixels) of a point p_1 in the left image and $p_2 = (x_2, y_2)$ represents the coordinates of the corresponding point in the right image in the table, the disparity is defined as

$$d(p_1, p_2) = \begin{cases} |x_1 - x_2|, & |y_1 - y_2| < \delta \\ 0, & \text{otherwise} \end{cases}$$

The disparity map assigns to each pixel of the left-camera image the disparity with its corresponding right image point (if one exists) from the table. If no corresponding pixel exists in the table, a value of 0 is assigned; δ in the above definition represents a user-assigned parameter. Appropriate values for δ remain to be experimentally determined.

The next stage of the data integration layer is to solve the stereo correspondence problem with the *a priori* disparity map. Details concerning the initial algorithm selected for the proof-of-concept experiments are provided in section 3 with results in section 4.

Once the stereo correspondence problem is solved, the final stage of the data integration layer (stereo 3-D reconstruction) is performed. A standard geometric triangulation algorithm is used. Details about the algorithm are given in Oberle and Haas (2002).

3. Application Domain and Stereo Correspondence Algorithm

3.1 Application Domain

As mentioned in the introduction, we are predominantly concerned with scenes representing complex environments. We define a complex environment as one in which there is a "large number" of depth discontinuities. Generally, this implies that the scene contains a relatively "large number" of individual objects at different depths. In addition, the individual objects will tend to be rather "thin" (e.g., trees or poles) and are called "narrow occluding objects" (Brown, Burschka, & Hager, 2003). An example of a complex environment is shown in figure 5.

In a stereo image pair, depth discontinuities result in occluded[9] points, i.e., scene elements visible in only one of the two images. This situation is illustrated in figure 6 where the portion of the object highlighted in red is visible in only the right camera. Since the stereo correspondence problem is already ill posed (Scharstein, Szeliski, & Zabih, 2001), occluded points only increase the difficulty of obtaining an accurate solution. Besides creating occluded points, narrow occluding objects create situations in which the "ordering constraint" is violated, further complicating solutions to the correspondence problem. Many of the algorithms developed to

[9] Also referred to as half-occluded points.

solve the correspondence problem at some point in their execution must choose between a number of potential correspondences (e.g., a pixel in the left image, depending on the criteria being used, may have several equally likely correspondences in the right image).

Figure 5. Example of a complex environment.

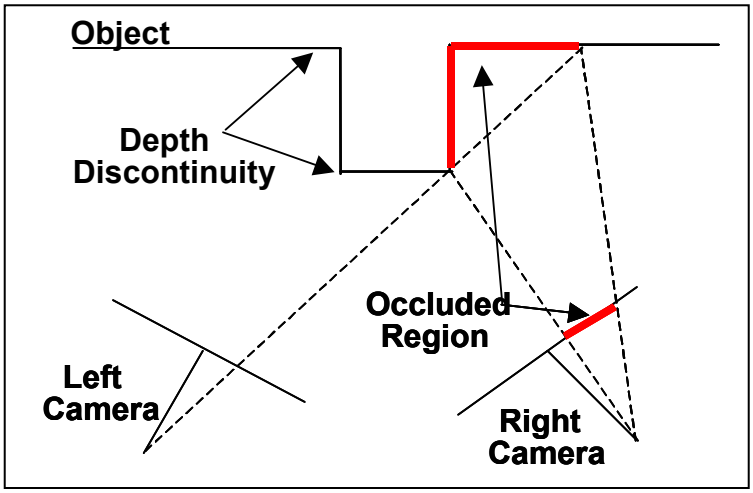

Figure 6. Example of occluded region.

To aid in the selection of the correct correspondence, the algorithms employ a number of global constraints. These constraints essentially represent prior knowledge concerning the scene. Several common constraints are smoothness of the disparity gradient, left-right consistency, and the ordering constraint. The ordering constraint basically assumes that image points will occur in the same order in both images. For an object with a continuous surface, this is true, even if the

surface is not at a constant depth relative to the cameras. This situation is illustrated in figure 7. However, if several distinct objects (especially if the objects are thin) are in the field of view of the cameras, the ordering constraint could fail, as shown in figure 8 (red and yellow points have switched order). See Dhond and Aggarwal (1992) for additional details involving the ordering constraint and stereo matching in the presence of thin occluding objects.

In summary, our desire to work with complex environments imposes two conditions on whatever algorithm is selected to solve the correspondence problem. First, the algorithm must be robust in terms of identifying occluded regions. The second condition is that the algorithm must not rely on the ordering constraint in resolving ties between possible matches.

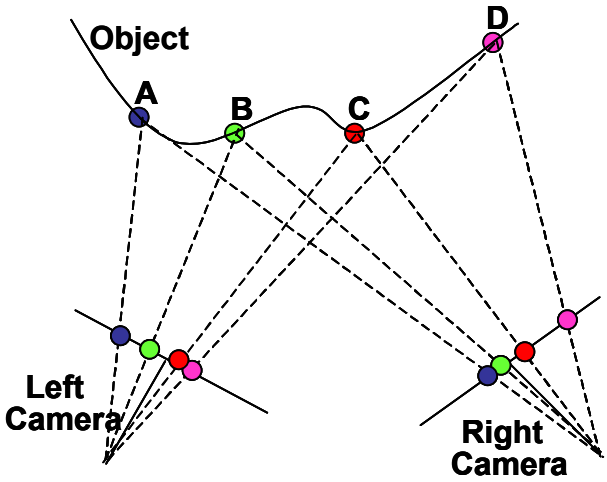

Figure 7. Example where ordering constraint holds.

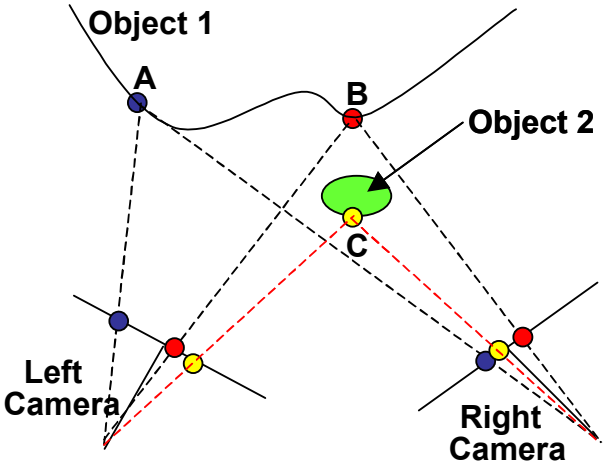

Figure 8. Example where ordering constraint fails.

3.2 Stereo Correspondence Algorithm

Stereo correspondence algorithms and their development is one of the most active research areas within the computer vision community. Thus, numerous stereo correspondence algorithms exist, which employ a variety of approaches available for use in our proof-of-concept experiments. In the end, as many as 30 different stereo correspondence algorithms were considered. Besides the two conditions stated before, other important considerations in the final algorithm selection are dense disparity maps[10], accuracy, and execution time. Fortunately, researchers at Middlebury College, Vermont (http://cat.middlebury.edu/stereo/; Scharstein & Szeliski, 2002; Scharstein, Szeliski, & Zabih, 2001) have maintained a web site over the past several years which contains stereo pairs (non-complex scenes with occlusions) with ground truth to permit the comparison of different stereo correspondence algorithms. The results compiled by the Middlebury College researchers are used in our final selection.

[10]A dense disparity map assigns to almost every pixel in one image a corresponding pixel in the other image or identifies the pixel as being occluded in the other image.

Following the taxonomy of Brown, Burschka, and Hager (2003), stereo correspondence algorithms are classified as local methods or global methods. Local methods base matching decisions on a small number of pixels surrounding a given pixel. For example, matching depends on intensity values within regularly sized neighborhoods of the pixels and some form of similarity (dis-similarity) measure, such as sum-of-squared differences or census metric (Banks & Corke, 2001; Sebe, Lew, & Huijsmans, 2000; Scherer, Werth, & Pinz, 1999; Bhat & Nayar, 1998), is used to establish the correspondences. Global methods base the matching decisions on scan lines or the entire image. Dynamic programming algorithms across scan lines and "graph cut" algorithms that determine the disparity map for the entire image simultaneously are examples of global methods.[11]

Local method algorithms are also referred to as window-, area-, or correlation-based algorithms. These algorithms represent some of the earliest developed to solve the stereo correspondence problem. Algorithms in this category tend to execute rapidly and form the basis for practically all "real-time" stereo implementations (Brown, Burschka, & Hager, 2003; Hirschmuller, 2001; Kimura, Shinbo, Yamaguchi, Kawamura, & Nakano, 1999). Although most of the local methods produce dense disparity maps, those methods based on matching features (e.g., occlusion edges, corners, or domain-specific features such as road surface markings) do not. On the other hand, feature-based methods tend to be less sensitive to depth discontinuities than other local methods. However, as Brown, Burschka, and Hager (2003) state, "Due to the need for dense depth maps for a variety of applications and also due to improvements in efficient and robust block-matching methods, interest in feature-based methods has declined in the last decade." Because of the lack of dense disparity maps, we rejected feature-matching algorithms for this work.

Two recent local method implementations that do not rely on the ordering constraint or feature-based matching have been submitted to the Middlebury College site for comparison with other methods. Muhlmann, Maier, Hesser, and Manner (2001) developed a correlation-based method using a median filtering to remove outliers and left-right consistency to eliminate false matches to generate a sub-pixel accurate disparity map. Although efficient, the algorithm is ranked approximately 27th of the 30 algorithms compared on the Middlebury College site. The most recent (April 15, 2004) results for the site are provided in appendix A. Hirschmuller (2001) also uses a correlation-based method. He uses a novel multiple window approach and a border correction filter to decrease matching errors at depth discontinuities. A general error filter is used to further invalidate uncertain matches. Although Hirschmuller's algorithm produces improved results for the Middlebury College comparisons, it still ranks approximately 17th. Based on these results, we made the decision not to use a local method.

[11]Other global methods, some of which perform rather well in the Middlebury College comparisons, include layered approaches (Baker, Szeliski, & Anandan, 1998; Shade, Gortler, He, & Szeliski, 1998), belief propagation (Sun, Zheng, & Shum, 2003), and Markov random fields (Boykov, Veksler, & Zabih,1998). Our analysis indicated that these approaches were not the most suitable for our work.

Our analysis of global methods indicated that the best choice to achieve our objectives is a graph-cut algorithm. Specific details concerning our choice are provided next. However, numerous other global methods exist and have been evaluated (e.g., see footnote 11). One method that is often used in stereo correspondence global methods is dynamic programming (Redert, Tsai, Hendriks, & Katsaggelos, 1998; Tsai & Katsaggelos, 1999), and we felt that several remarks about this method and why it was not selected are warranted. Cormen, Leiserson, and Rivest (1990) define dynamic programming as a mathematical method that reduces the computational complexity of an optimization problem by decomposing it into smaller and simpler sub-problems. Thus, dynamic programming is not specific to stereovision. A global cost function across scan lines is computed in stages. Going from one stage to the next is determined by a set of constraints. One of the necessary constraints is the ordering constraint (Amini, Weymouth, & Jain, 1990). Since one of our conditions for the stereo correspondence algorithm is that it cannot depend on the ordering constraint, no algorithm using dynamic programming is acceptable for our work.

Starting in the mid-1990's, a new global method approach to the stereo correspondence problem was developed, based on the minimization of an "energy function" using graph cuts (Boykov, Veksler, & Zabih, 2001(A), 2001(B); Boykov & Kolmogorov, 2001; Kolmogorov & Zabih, 2001; Kolmogorov & Zabih, 2002; Kolmogorov, Zabih, & Gortler, 2003). Minimization of an energy function is well suited to our situation. It is reasonable to expect that the solution to our correspondence problem will not vary far from the *a priori* ladar disparity information. In addition, graph-cut algorithms do not require the use of the ordering constraint. Thus, we chose to use a graph-cut methodology for solving the stereo correspondence problem.

Kolmogorov and Zabih (2002) describe the graph-cut approach as "The basic technique is to construct a specialized graph for the energy function to be minimized, such (sic) that the minimum cut on the graph also minimizes the energy (either globally or locally). The minimum cut in turn can be computed very efficiently by max (sic) flow algorithms." Unfortunately, as they state, "Minimizing an energy function via graph cuts, however, remains a technically difficult problem. Each paper constructs its own graph specifically for its individual energy function, and in some of these cases, the construction is fairly complex." Since our goal is directed toward data integration and fusion, not the development of a graph-cut algorithm, we elected to modify an existing graph-cut algorithm. An algorithm by Kolmogorov and Zabih (2001) described in *Computing Visual Correspondence with Occlusions Using Graph Cuts* was selected for its explicit handling of occlusions. Code for their algorithm is available on line at http://www.cs.cornell.edu/People/vnk/software.html.

The basic steps for using a graph-cut algorithm are

1. Define the energy function,
2. Construct the appropriate graph, and
3. Use a maximum flow algorithm to minimize the energy function via graph cuts.

As mentioned, the algorithm used in our work is a modification of the Kolmogorov and Zabih (2001) algorithm. Our modification is the inclusion of an additional term in the energy function representing the cost or penalty of assigning a disparity to a pixel different from that assigned by the *a priori* ladar disparity data. As long as this term is non-negative, the results of their algorithm (i.e., computation of a strong local minimum[12] for the energy function) remain valid (Kolmogorov & Zabih, 2002).

A brief description of the modified energy function using the notation of Kolmogorov and Zabih (2001) is provided. This illustrates our modification of the original Kolmogorov and Zabih energy function. For details concerning the construction of the appropriate graph and the use of a new maximum flow algorithm based on α-expansion (Boykov & Kolmogorov, 2001) to minimize the energy function, the reader is referred to Kolmogorov and Zabih (2001).

Notation:

$\boldsymbol{P} =$ set of all pixels, i.e., pixels left image $\cup$ pixels right image.

$\boldsymbol{A} = \{\langle p,q\rangle \mid p \text{ and } q \text{ are pixels in different images}\}$,

i.e., a set of unordered pairs of pixels that could potentially correspond. An element of $\boldsymbol{A}$ is termed an "assignment."

$\mathrm{d}(\langle p,q\rangle) =$ disparity between pixels p and q.

$\boldsymbol{f}$: assigns a 1 or 0 to every element (assignment) of $\boldsymbol{A}$, referred to as a "configuration." An assignment of $\boldsymbol{A}$ is termed active if it is assigned a value of 1. Active assignments can be thought of as pixels that correspond.

$\boldsymbol{A}(\boldsymbol{f}) =$ subset of $\boldsymbol{A}$ consisting of active assignments according to the configuration $\boldsymbol{f}$.

$\boldsymbol{N}_p(\boldsymbol{f}) = \{\langle p,q\rangle \in \boldsymbol{A}(\boldsymbol{f})\}$, set of active assignments in $\boldsymbol{f}$ that involve pixel p.

Unique configuration $\boldsymbol{f}$: $\forall p \in \boldsymbol{P} \quad |\boldsymbol{N}_p(\boldsymbol{f})| \le 1$, i.e., each pixel is involved in one active assignment at most. Note that occluded pixels satisfy $|\boldsymbol{N}_p(\boldsymbol{f})| = 0$.

$$\boldsymbol{T}(\cdot) = \begin{cases} 1, & \text{argument true or non-zero} \\ 0, & \text{otherwise} \end{cases}.$$

$$\mathcal{N} = \left\{ \{a1,a2\} \middle| a1 \in \boldsymbol{A}, a2 \in \boldsymbol{A}, \mathrm{d}(a1) = \mathrm{d}(a2), \text{ and if } a1 = \langle p,q\rangle \text{ and } a2 = \langle r,s\rangle \text{ with } p \text{ and } r \right. \\ \left. \text{in the left image, then } p \text{ and } r \text{ or } q \text{ and } s \text{ are adjacent pixels} \right\}$$

[12] Within a known factor of the global minimum

Energy Function:

Employing the previous notation, our modification of the Kolmogorov and Zabih (2001) energy function is written in general as

$$\mathrm{E}(\boldsymbol{f}) = \mathrm{E}_{\mathrm{data}}(\boldsymbol{f}) + \mathrm{E}_{\mathrm{ladar}}(\boldsymbol{f}) + \mathrm{E}_{\mathrm{occ}}(\boldsymbol{f}) + \mathrm{E}_{\mathrm{smooth}}(\boldsymbol{f}),$$

with our modification being the $\mathrm{E}_{\mathrm{ladar}}(\boldsymbol{f})$ term. The data term is the cost associated with an assignment being identified as active and is given by

$$\mathrm{E}_{\mathrm{data}}(\boldsymbol{f}) = \sum_{a \in \boldsymbol{A}(\boldsymbol{f})} \mathrm{D}(a),$$

in which for an assignment $a = \langle p, q \rangle, \mathrm{D}(a) = \left(I(p) - I(q)\right)^2$, with I the intensity of the pixel. The ladar term is the cost associated with an active assignment that has a disparity different from the *a priori* ladar disparity data and is given by

$$\mathrm{E}_{\mathrm{ladar}}(\boldsymbol{f}) = \sum_{a \in \boldsymbol{A}(\boldsymbol{f})} \left(\mathrm{d}(a) - \mathcal{L}(p)\right)^2 \cdot \boldsymbol{T}\left(\mathcal{L}(p)\right).$$

In the expression, p is the left-image pixel of the assignment a, and $\mathcal{L}(p)$ is the *a priori* ladar disparity assigned to pixel p. Note that if no disparity is assigned to the pixel from the ladar information, $\mathcal{L}(p) = 0$ and no cost is incurred. Thus, the *a priori* ladar data only influence those pixels that are in the range of the projection of the ladar data onto the left and right camera images. In addition, if the resolution of the ladar data is low, $\mathcal{L}(p)$ will equal 0 for most active assignments. A major research effort of this work is to investigate the effect on the solution to the correspondence problem resulting from different approaches for extending the *a priori* disparity information to all active assignments. The occlusion term imposes a cost for identifying a pixel as occluded. This term is given by

$$\mathrm{E}_{\mathrm{occ}}(\boldsymbol{f}) = \sum_{p \in \boldsymbol{P}} C_p \cdot \boldsymbol{T}\left(\left|\boldsymbol{N}_p(\boldsymbol{f})\right| = 0\right).$$

The value of C_p is defined next. Finally, the smoothness term imposes a cost if adjacent pixels in the same image do not have the same disparity. In terms of assignments, this is equivalent to imposing a cost if one assignment is present in the configuration and another close assignment with the same disparity is not. Specifically the smoothness term is given by

$$\mathrm{E}_{smooth}(\boldsymbol{f}) = \sum_{\{a1,a2\} \in \mathcal{N}} V_{a1,a2} \cdot \boldsymbol{T}\left(\boldsymbol{f}(a1) \neq \boldsymbol{f}(a2)\right).$$

Details of the function $V_{a1,a2}$ are provided next.

The goal is to determine the unique configuration, $\boldsymbol{f}^*$, that minimizes $E(\boldsymbol{f})$. The solution to the stereo correspondence problem follows from the minimizing configuration. Active assignments

identify corresponding pixels from which the disparity can be determined, while all pixels not included in an active assignment are classified as occluded.

To complete the description of the energy function C_p and $V_{a1,a2}$ must be defined. Let $a1 = \langle p, q \rangle$ and $a2 = \langle r, s \rangle$ be two assignments with p and r in the same image. C_p and $V_{a1,a2}$ are then defined as

$$C_p = \lambda,$$

and

$$V_{a1,a2} = \begin{cases} \lambda & \text{if } \max\left(\left|I(p) - I(r)\right|, \left|I(q) - I(s)\right|\right) < 8 \\ 3\lambda & \text{otherwise} \end{cases}.$$

The value of λ is chosen empirically, or in the case of the Kolmogorov and Zabih implementation that we use, λ can also be automatically determined. We chose to allow the code to automatically determine λ, since results presented by Kolmogorov and Zabih (2001) indicated that their method is relatively insensitive to the specific choice of λ.

4. Proof-of-Concept Calculations

Since we do not have simultaneous ladar and stereo camera data supported by ground truth for scenes from complex environments, the ground truth imagery from the University of Tsukuba, Japan (Scharstein & Szeliski, 2002, 2003) is used. The Tsukuba imagery used is the left and right camera images (figure 9) and the ground truth images of disparity and occluded pixels (figure 10).[13]

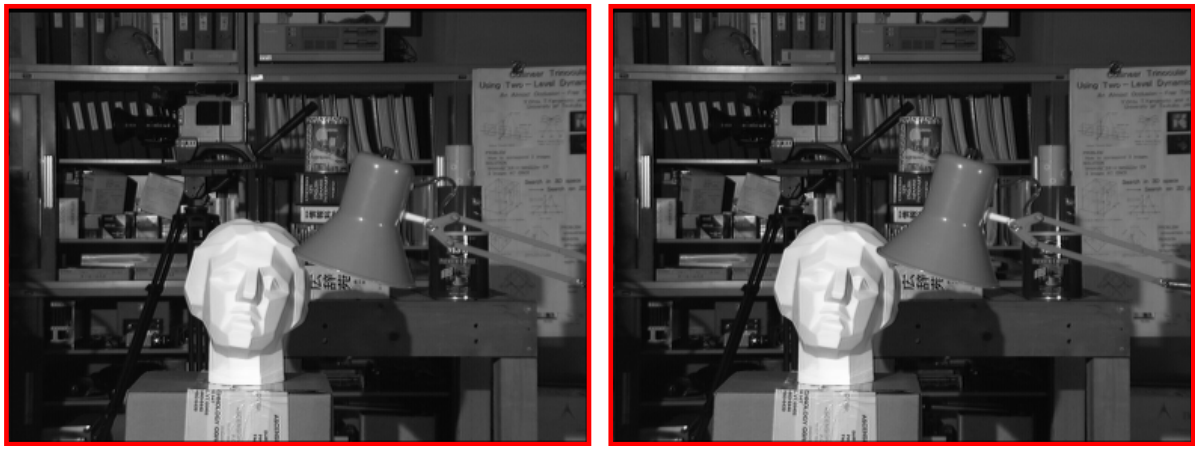

Figure 9. Left (left) and right (right) Tsukuba stereo images.

[13]The University of Tsukuba imagery set is available on line at the Middlebury College Vision web site. Y. Ohta and Y. Nakamura of the University of Tsukuba supplied the imagery data set.

The images are 384 by 288 pixels. The ground truth images (figure 10) have a border of 18 pixels in which there is no information. In the disparity image (left side of figure 10), this is the black border.

For the proof-of-concept calculations, the *a priori* ladar disparity data are taken to be the ground truth disparity data (left side of figure 10). Some differences exist between the ground truth disparity data and what would be expected from actual *a priori* ladar disparity data. The 18-pixel border of the ground truth disparity data will be incorrectly interpreted as occluded when treated as the *a priori* ladar data. In addition, the ground truth disparity data has been "filled in" so that all pixels are assigned a disparity even if the pixel is actually occluded (i.e., the pixels identified as occluded in the image in the right of figure 10 are assigned disparities in the image in the left of figure 10). Results of calculations with and without the use of the *a priori* ladar disparity data are presented in table 1. Comparisons are relative to the ground truth information and the calculated results.

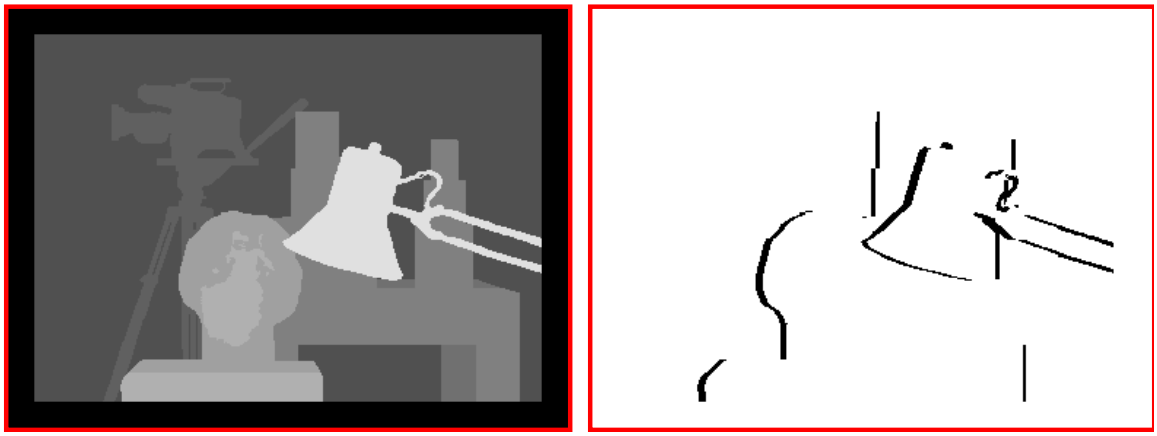

Figure 10. Ground truth for Tsukuba imagery, disparity (left) and occlusions (right).

Table 1. Results of proof-of-concept calculations.

	Percentage of Pixels Whose Disparity Correctly Labeled (pixels labeled as occluded by both the calculation and ground truth ignored)	**Percentage of Occluded Pixels Correctly Labeled**	**Percentage of Pixels Incorrectly Labeled as Occluded**
Calculation with *a priori* ladar disparity data	92.955	77.4	1.40
Calculation without *a priori* ladar disparity data	92.343	67.6	1.45

Based on the results of the Middlebury College comparisons (appendix A), the Kolmogorov and Zabih algorithm is either 1 or 2 in terms of its performance for the Tsukuba image pair. Thus, the small increase in the percentage of correctly labeled pixels using the *a priori* ladar disparity data is encouraging. More encouraging is the improvement in the results related to occluded pixels for the calculation using the *a priori* ladar disparity data. As discussed before, the *a priori*

ladar disparity data used in the calculation provided erroneous occlusion information, yet the calculation with the *a priori* ladar disparity data correctly identified approximately 15% (77.4% versus 67.6%) more occluded pixels compared to the calculation without the *a priori* ladar disparity data. Those pixels identified as occluded in both calculations are shown in figure 11. Results for the calculation with the *a priori* ladar disparity data are shown on the left of figure 11, and results for the calculation without the *a priori* ladar disparity data are shown on the right. The results for the calculation with the *a priori* ladar disparity data (left side of figure 11) appear for the most part to be cleaner (e.g., areas inside red circles) than for the results without the *a priori* ladar disparity data (right side of figure 11).

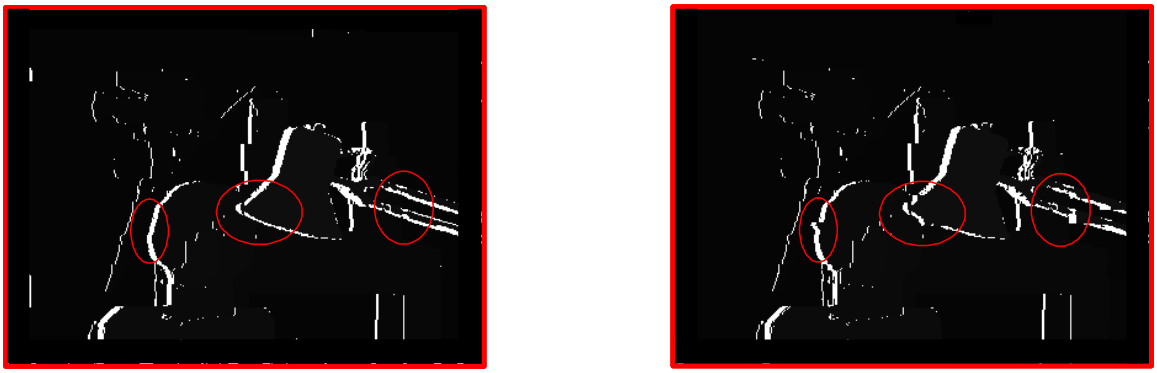

Figure 11. Occluded pixels for calculation with (left) and without (right) *a priori* ladar disparity data.

In addition, the *a priori* ladar disparity data calculation incorrectly labeled roughly 3.5% (1.40% versus 1.45%) fewer pixels as occluded compared to the other calculation.

Improved results for the calculation with the *a priori* ladar disparity data are also evident in an analysis of the pixels that were incorrectly labeled. The maximum disparity that could be assigned to a pixel for the calculations is 15. For both calculations (with and without *a priori* ladar disparity data) the largest difference between any calculated disparity and the ground truth disparity is 13. As illustrated in figure 12, the errors in the disparity assignments for the calculation with *a priori* ladar disparity data are generally smaller, compared to the calculation without *a priori* ladar disparity data, with no disparity error greater than 6 compared to 13 for the other calculation.

Finally, a series of calculations was performed in which "white noise" was added to the left and right stereo images. As expected, the results degraded with the difference between the calculations using the *a priori* ladar disparity data increasing as the severity of the noise increased. Details of these calculations are not provided.

Based on the overall results for the different calculations, it appears that improvements in the solution of the stereo correspondence problem can result from the use of *a priori* ladar disparity data. Improvements relative to the solution without the *a priori* ladar disparity data are observed in the number of correctly labeled pixels (disparity and occlusions) and a reduction in the magnitude of the error in the disparity for those pixels that are incorrectly labeled.

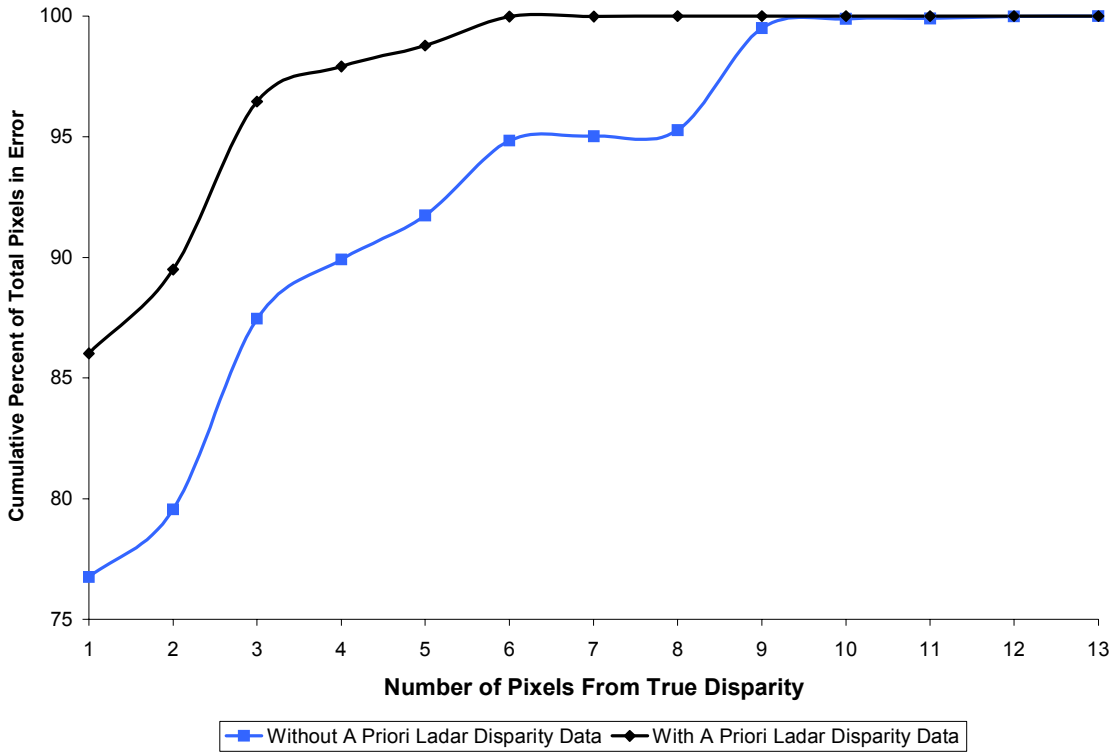

Figure 12. Cumulative percentage of total pixels in error versus error in disparity.

5. Summary and Future Work

In this report, we described an approach and architecture to incorporate data integration and fusion of ladar sensor data and stereo camera imagery to produce high resolution, ladar-quality 3-D world models. Of particular interest is the construction of world models for scenes involving complex environments—a situation that is extremely difficult for traditional stereo algorithms because of the large number of occluded regions. As stated earlier in the report, the principal novelty of our work is the integration of 3-D ladar information as an *a priori* disparity map to improve the solution to the stereo correspondence problem. Proof-of-concept[14] calculations were performed with a modified energy function based upon the work of Kolmogorov and Zabih. The approach uses recently developed algorithms for computer vision incorporating minimum cut-maximum flow paradigms. Our results indicated that data integration of ladar data as an *a priori* disparity map and stereo data can produce improvements in the solution to the correspondence problem. Of particular note is the improvement in the identification of occluded regions with the data integration.

[14]Data integration can improve the solution to the stereo correspondence problem.

Although we describe a detailed architecture, a number of the necessary algorithms in the data integration and fusion layers have not been developed. This is especially true for the data fusion layer. However, in this layer, we expect to draw heavily on the many efforts involving data fusion described in the vision literature. Near-term future work will be directed toward completing the algorithms of the data integration layer. Hopefully, this will include an evaluation of additional stereo correspondence algorithms specifically developed for complex environments.

6. References

Abidi, M. A.; Gonzalez, R. C., Ed. *Data Fusion in Robotics and Machine Intelligence*, Academic Press, 1992.

Amini, A.; Weymouth, T.; Jain, R. Using Dynamic Programming for Solving Variational Problems in Vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **September 1990**, *12* (9), 855–867.

Baker, S.; Szeliski, R.; Anandan, P. A Layered Approach to Stereo Reconstruction. *Proceedings 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1998, 434–441, June 1998.

Banks, J.; Corke, P. Quantitative Evaluation of Matching Methods and Validity Measures for Stereo Vision. *International Journal of Robotics Research* **July 2001**, *20* (7), 512–532.

Bhat, D. N.; Nayar S. K. Ordinal Measures for Image Correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **April 1998**, *20* (4), 415–423.

Bischoff, R.; Graefe, V. Machine Vision for Intelligent Robots. *International Association for Pattern Recognition (IAPR) Workshop on Machine Vision Applications* **November 1998**, 167–176, Makuhari/Tokyo.

Bouguet, J.-Y. *Camera Calibration Toolbox for Matlab*, web site: www.vision,caltech.edu/bouguetj/calib.doc, 2003.

Boykov, Y.; Kolmogorov, V. An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition* **September 2001**, 359–374.

Boykov, Y.; Veksler, O.; Zabih, R. Markov Random Fields with Efficient Approximations. *IEEE Conference on Computer Vision and Pattern Recognition* **June 1998**, 648–655.

Boykov, Y.; Veksler, O.; Zabih, R. V. A New Algorithm for Energy Minimization with Discontinuities. *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition* **2001(A)**, 205–220.

Boykov, Y.; Veksler, O.; Zabih, R. Fast Approximate Energy Minimization via Graph Cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **November 2001(B)**, *23* (11), 1222–1239.

Brown, M. Z.; Burschka, D.; Hager, G. D. Advances in Computational Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* **August 2003**, *25* (8), 993–1008.

Chang, T.; Hong, T. H.; Rasmussen, C. E.; Shneier, M. O. Fusing Ladar and Color Image Information for Mobile Robot Feature Detection and Tracking. *7th International Conference on Intelligent Autonomous Systems (IAS-7)*, CA, March 2002.

Cormen, T. H.; Leiserson, C. E.; Rivest, R. L. *Introduction to Algorithms*; MIT Press, McGraw-Hill, New York, New York, 1990.

DeMenthon, D. F.; Davis, L. S. Model-Based Object Pose in 25 Lines of Code. *International Journal of Computer Vision* **1995**, *15*, 123–141.

Dhond, U. R.; Aggarwal, J. K. Computing Stereo Correspondence in the Presence of Narrow Occluding Objects. *1992 IEEE Proceeding on Computer Vision and Pattern Recognition (CVPR92)*, 758–760, 1992.

Dias, P.; Sequeira, V.; Goncalves, J.G.M.; Vaz, F. Fusion of intensity and range data for improved 3D models. *IEEE Proceedings 2001 International Conference on Image Processing* **7-10 October 2001**, *3*, 1107–1110.

Elstrom, M. D. *A Stereo-Based Technique for the Registration of Color and LADAR Images*, Master's Thesis, The University of Tennessee, Knoxville, August 1998. web site: imaging.utk.edu/publications/papers/ dissertation/elstrom.pdf.

Faugeras, O. *Three-Dimensional Computer Vision A Geometric Viewpoint*; The MIT Press, Cambridge, Massachusetts, 1993.

Hall, D. L. *Mathematical Techniques in Multisensor Data Fusion*; ARTECH House, Inc., Norwood, MA, 1992.

Hirschmuller, H. Improvements in Real-time Correlation-based Stereo Vision. *IEEE Workshop on Stereo and Multi-Baseline Vision* **December 2001**, 141–148.

Kimura, S.; Shinbo, T.; Yamaguchi, H.; Kawamura, E.; Nakano, K. A Convolver-based Real-time Stereo Machine (SAZAN). *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **June 1999**, *1*, 457–463.

Kolmogorov, V.; Zabih, R. Computing Visual Correspondence with Occlusions Using Graph Cuts. *Proceedings Eighth IEEE International Conference on Computer Vision* **July 2001**, *II*, 508–515.

Kolmogorov, V.; Zabih, R. What Energy Functions can be Minimized via Graph Cuts. *European Conference on Computer Vision* **2002**, *3*, 65–81.

Kolmogorov, V.; Zabih, R.; Gortler, S. Generalized Multi-camera Scene Reconstruction Using Graph Cuts. *Fourth International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, July 2003.

Lim, S.-N.; Mittal, A.; Davis, L. S. *Uncalibrated Stereo Rectification for Automatic 3D Surveillance*, Manuscript in review, University of Maryland, College Park, Maryland, 2004.

Loop, C.; Zhang, Z. Computing Rectifying Homographies for Stereo Vision. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **June 1999**, *I*, 125–131.

Muhlmann, K.; Maier, D.; Hesser, J.; Manner, R. Calculating Dense Disparity Maps from Color Stereo Images, an Efficient Implementation. *Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision*, 30–36, December 2001.

Nickels, K. M.; Castano, A.; Cianci, C. Fusion of Lidar and Stereo Range for Mobile Robots. *Proceedings of ICAR 2003*, 65–70, The 11th International Conference on Advanced Robotics, Coimbra, Portugal, 30 June – 3 July, 2003.

Oberle, W. F. *The Effect of Variability in Stereo Camera Calibration and Registration Parameters on Three-Dimensional Reconstruction Distance Estimates*; ARL-TR-3140; U. S. Army Research Laboratory: Aberdeen Proving Ground, MD, February 2004.

Oberle, W. F.; Haas, G. A. *Three-Dimensional Stereo Reconstruction and Sensor Registration With Application to the Development of a Multi-Sensor Database*; ARL-TR-2878; U. S. Army Research Laboratory: Aberdeen Proving Ground, MD, December 2002.

Oberle, W. F.; Haas, G. A. *LADAR – Camera Registration Using Object Pose and Camera Calibration Algorithms*; ARL-TR-3147; U. S. Army Research Laboratory: Aberdeen Proving Ground, MD, April 2004.

Pollefeys, M.; Koch, R.; Van Gool, L. A Simple and Efficient Rectification Method for General Motion. *Proceedings of International Conference on Computer Visions (ICCV'99)*, Vol. 1, 496–501, Corfu, Greece, September 1999.

Redert, A.; Tsai, C.-J.; Hendriks, E.; Katsaggelos, A. K. Disparity Estimation with Modeling of Occlusion and Object Orientation. *Proceedings of the SPIE Visual Communications and Image Processing*, Volume 3309, 798–808, January 1998.

Scharstein, D.; Szeliski, R. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision*, **April–June 2002**, *47*(1/2/3), 7–42.

Scharstein, D.; Szeliski, R. High Accuracy Stereo Depth Maps Using Structural Light. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR 2003), Volume 1, pp 195-202, Madison, WI, June 2003.

Scharstein, D.;Szeliski, R.;Zabih, R. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *IEEE Workshop on Stereo and Multi-Baseline Vision*, 131–140, December 2001.

Scherer S.; Werth P.; Pinz, A. The Discriminatory Power of Ordinal Measures – Towards a New Coefficient. *IEEE Computer Science Conference on Computer Vision and Pattern Recognition*, Volume 1, 76–81, June 1999.

Sebe, N.; Lew, M. S.; Huijsmans, D. P. Toward Improved Ranking Metrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **October 2000**, *22* (10), 1132–1143.

Shade, J.; Gortler, S.; He, L.-W.; Szeliski, R. Layered Depth Images. *SIGGRAPH 98, Computer Graphics Proceedings*, 231–242, July 1998.

Spero, D. J.; Jarvis, R. A. 3D Vision for Large-Scale Outdoor Environments. *Proceedings 2002 Australasian Conference on Robotics and Automation*, Auckland, 27 – 29 November 2002.

Stevens, M. R.; Beveridge, J. R.; Goss, M. E. *Visualizing Multisersor Model-Based Object Recognition*, Reconnaissance, Surveillance, and Target Acquisition for the Unmanned Ground Vehicle: Providing Surveillance "Eyes" for an Autonomous Vehicle, Editors: Oscar Firschein and Thomas M. Strat, Morgan Kaufmann Publishers, ISBN 1-55860-451-0, 1997.

Sun, J.; Zheng, N. N.; Shum, H. Y. Stereo Matching Using Belief Propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **July 2003**, *25* (7), 787–800.

Trucco, E.; Verri, A. *Introductory Techniques for 3-D Computer Vision*; Prentice Hall, Upper Saddle River, New Jersey, 07458, 1998.

Tsai, C.-J.; Katsaggelos, A. K. Dense Disparity Estimation with a Divide-and-Conquer Disparity Space Image Technique. *IEEE Transactions on Multimedia* **March 1999**, *1*, 18–29.

INTENTIONALLY LEFT BLANK

Appendix A. Results from Middlebury College Stereo Vision Comparison (April 15, 2004) (http://cat.middlebury.edu.stereo/)

Welcome to the Middlebury Stereo Vision Page

This web site contains material accompanying our taxonomy and experimental comparison of stereo correspondence algorithms [1]. It contains stereo data sets with ground truth, the overall comparison of algorithms, instructions on how to evaluate your stereo algorithm in our framework, and our stereo correspondence software.

Also available are two new stereo data sets with ground truth obtained using our structured lighting technique [2]. These data sets have a more complex geometry and larger disparity ranges than the original data sets.

We are continually inviting other researchers to run their stereo algorithms on the four image pairs used in our overall comparison, and to send us the results. We will then run our evaluator, and report the resulting disparity error statistics. If you are interested in participating, please go to the evaluation page.

How to Cite the Materials on This Web Site:

We grant permission to use and publish all images and numerical results on this website. However, if you use our data sets, and/or report performance results, we request that you cite the appropriate paper(s) [1, 2]. If you want to cite this website, please use the "stable" URL "**www.middlebury.edu/stereo**". (This URL is currently auto-forwarded to "cat.middlebury.edu/stereo", but that may change.)

References:

[1] D. Scharstein and R. Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms.
IJCV 47(1/2/3):7-42, April-June 2002. PDF file (1.15 MB) - includes current evaluation.
Microsoft Research Technical Report MSR-TR-2001-81, November 2001. PDF file (1.27 MB).

[2] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light.
In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003),* volume 1, pages 195-202, Madison, WI, June 2003. PDF file (1.2 MB)

Support for this work was provided in part by NSF CAREER grant 9984485. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Table A-1. Comparison of the performance of different stereo algorithms on four test image pairs

Algorithm	**Tsukuba**			**Sawtooth**			**Venus**			**Map**	
	all	untex.	disc.	**all**	untex.	disc.	**all**	untex.	disc.	**all**	disc.
Segm.-based GC [23]	1.23 3	0.29 2	6.94 4	0.30 3	**0.00** 1	3.24 3	**0.08** 1	**0.01** 1	**1.39** 1	1.49 19	15.46 24
Segm.+glob.vis. [25]	1.30 5	0.48 5	7.50 6	**0.20** 1	**0.00** 1	**2.30** 1	0.79 4	0.81 5	6.37 7	1.63 21	16.07 26
Layered [16]	1.58 7	1.06 9	8.82 8	0.34 4	**0.00** 1	3.35 4	1.52 10	2.96 19	2.62 3	0.37 10	5.24 10
Belief prop. [3]	**1.15** 1	0.42 3	**6.31** 1	0.98 10	0.30 14	4.83 8	1.00 6	0.76 4	9.13 13	0.84 16	5.27 11
MultiCam GC [21]	1.85 10	1.94 15	6.99 5	0.62 8	**0.00** 1	6.86 12	1.21 8	1.96 10	5.71 6	0.31 7	4.34 9
Region-Progress. [24]	1.44 6	0.55 6	8.18 7	0.24 2	**0.00** 1	2.64 2	0.99 5	1.37 8	6.40 8	1.49 20	17.11 27
GC+occl. [2b]	1.19 2	**0.23** 1	6.71 2	0.73 9	0.11 9	5.71 10	1.64 13	2.75 17	5.41 5	0.61 13	6.05 12
Improved Coop. [19]	1.67 8	0.77 7	9.67 11	1.21 13	0.17 12	6.90 13	1.04 7	1.07 6	13.68 18	0.29 5	3.65 6
GC+occl. [2a]	1.27 4	0.43 4	6.90 3	0.36 5	**0.00** 1	3.65 5	2.79 21	5.39 22	2.54 2	1.79 22	10.08 18
Disc. pres. [18]	1.78 9	1.22 11	9.71 12	1.17 12	0.08 8	5.55 9	1.61 12	2.25 13	9.06 12	0.32 8	3.33 5
Symbiotic [20]	2.87 15	1.71 14	11.90 13	1.04 11	0.13 10	7.32 15	0.51 2	0.23 2	7.88 10	0.50 12	6.54 13
Graph cuts [1a]	1.94 12	1.09 10	9.49 10	1.30 15	0.06 7	6.34 11	1.79 16	2.61 16	6.91 9	0.31 6	3.88 7
Var. win. [17]	2.35 13	1.65 13	12.17 15	1.28 14	0.23 13	7.09 14	1.23 9	1.16 7	13.35 17	0.24 3	2.98 3
Graph cuts [5]	1.86 11	1.00 8	9.35 9	0.42 6	0.14 11	3.76 6	1.69 15	2.30 14	5.40 4	2.39 25	9.35 16
Multiw. cut [13]	8.08 27	6.53 24	25.33 28	0.61 7	0.46 17	4.60 7	0.53 3	0.31 3	8.06 11	0.26 4	3.27 4
Comp. win. [4]	3.36 18	3.54 18	12.91 18	1.61 18	0.45 16	7.87 16	1.67 14	2.18 11	13.24 16	0.33 9	3.94 8
Realtime [7]	4.25 22	4.47 22	15.05 22	1.32 16	0.35 15	9.21 17	1.53 11	1.80 9	12.33 14	0.81 15	11.35 21
Cooperative [6]	3.49 19	3.65 19	14.77 20	2.03 19	2.29 23	13.41 22	2.57 20	3.52 20	26.38 27	0.22 2	**2.37** 1
Bay. diff. [1b]	6.49 26	11.62 29	12.29 16	1.45 17	0.72 18	9.29 18	4.00 23	7.21 25	18.39 22	**0.20** 1	2.49 2
Stoch. diff. [9]	3.95 20	4.08 21	15.49 24	2.45 23	0.90 20	10.58 19	2.45 18	2.41 15	21.84 24	1.31 18	7.79 15
Genetic [11]	2.96 16	2.66 17	14.97 21	2.21 21	2.76 25	13.96 23	2.49 19	2.89 18	23.04 25	1.04 17	10.91 20
SSD+MF [1c]	5.23 25	3.80 20	24.66 27	2.21 20	0.72 19	13.97 24	3.74 22	6.82 24	12.94 15	0.66 14	9.35 16
Max flow [14]	2.98 17	2.00 16	15.10 23	3.47 24	3.00 26	14.19 25	2.16 17	2.24 12	21.73 23	3.13 26	15.98 25
Pix-to-pix [12]	5.12 24	7.06 27	14.62 19	2.31 22	1.79 21	14.93 26	6.30 26	11.37 28	14.57 19	0.50 11	6.83 14
Scanl. opt. [1e]	5.08 23	6.78 25	11.94 14	4.06 25	2.64 24	11.90 20	9.44 29	14.59 29	18.20 21	1.84 23	10.22 19
Dyn. prog. [1d]	4.12 21	4.63 23	12.34 17	4.84 28	3.71 28	13.26 21	10.10 30	15.01 30	17.12 20	3.33 27	14.04 23
Realtime DP [26]	2.85 14	1.33 12	15.62 25	6.25 30	3.98 29	25.19 28	6.42 27	8.14 26	25.30 26	6.45 29	25.16 28
MMHM [15]	9.76 29	13.85 30	24.39 26	4.76 27	1.87 22	22.49 27	6.48 28	10.36 27	31.29 28	8.42 30	12.68 22
Shao [8]	9.67 28	7.04 26	35.63 29	4.25 26	3.19 27	30.14 30	6.01 25	6.70 23	43.91 30	2.36 24	33.01 30
Max. surf. [10]	11.10 30	10.70 28	41.99 30	5.51 29	5.56 30	27.39 29	4.36 24	4.78 21	41.13 29	4.17 28	27.88 29

Our implementation:

[1] D. Scharstein and R. Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms, IJCV, 2002. Five algorithm have been implemented:
a - Graph cuts using alpha-beta swaps (Boykov, Veksler, and Zabih, PAMI 2001);
b - Bayesian diffusion (Scharstein and Szeliski, IJCV 1998);
c - SSD + min-filter (i.e., shiftable windows), window size = 21;
d - Dynamic programming, similar to Bobick and Intille (IJCV 1999);
e - Scanline optimization (1D optimization using horizontal smoothness terms).

Other authors' implementations:

[2] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions using graph cuts. ICCV 2001.
a - original submission
b - new submission with automatic parameter setting (same as in [21])

[3] J. Sun, H. Y. Shum, and N. N. Zheng. Stereo matching using belief propagation. PAMI 2003 (also in ECCV 2002).

[4] O. Veksler. Stereo matching by compact windows via minimum ratio cycle. ICCV 2001.

[5] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. PAMI 2001.

[6] L. Zitnick and T. Kanade. A cooperative algorithm for stereo matching and occlusion detection. PAMI 2000.

[7] H. Hirschmüller. Improvements in Real-Time Correlation-Based Stereo Vision. CVPR 2001 Stereo Workshop / IJCV 2002.

[8] J. Shao. Combination of Stereo, Motion and Rendering for 3D Footage Display. CVPR 2001 Stereo Workshop / IJCV 2002.

[9] S. H. Lee, Y. Kanatsugu, and J.-I. Park. Hierarchical stochastic diffusion for disparity estimation. CVPR 2001 Stereo Workshop / IJCV 2002.

[10] C. Sun. Fast stereo matching using rectangular subregioning and 3D maximum-surface techniques. CVPR 2001 Stereo Workshop / IJCV 2002.

[11] M. Gong and Y.-H. Yang. Multi-baseline Stereo Matching Using Genetic Algorithm. CVPR 2001 Stereo Workshop / IJCV 2002.

[12] S. Birchfield and C. Tomasi. Depth discontinuities by pixel-to-pixel stereo. ICCV 1998.

[13] S. Birchfield and C. Tomasi. Multiway cut for stereo and motion with slanted surfaces. ICCV 1999.

[14] S. Roy and I. J. Cox. A maximum-flow formulation of the N-camera stereo correspondence problem. ICCV 1998.

[15] K. Mühlmann, D. Maier, J. Hesser, and R. Männer. Calculating Dense Disparity Maps from Color Stereo Images, an Efficient Implementation. CVPR 2001 Stereo Workshop / IJCV 2002.

[16] M. Lin and C. Tomasi. Surfaces with Occlusions from Layered Stereo. Ph.D. thesis, Stanford University, 2002.

[17] O. Veksler. Fast Variable Window for Stereo Correspondence using Integral Images. CVPR 2003.

[18] M. Agrawal and L. Davis. Window Based, Discontinuity Preserving Stereo. Submitted to CVPR 2003.

[19] H. Mayer. Analysis of Means to Improve Cooperative Disparity Estimation. ISPRS Conf. on Photogrammetric Image Analysis, 2003.

[20] J. Y. Goulermas and P. Liatsis. A Collective-based Adaptive Symbiotic Model for Surface Reconstruction in Area-based Stereo. IEEE Trans. Evolutionary Computation, vol.7(5), pp.482-502, 2003.

[21] V. Kolmogorov and R. Zabih. Multi-camera Scene Reconstruction via Graph Cuts. ECCV 2002.

[22] (Withdrawn)

[23] L. Hong and G. Chen. Segment-Based Stereo Matching Using Graph Cuts. CVPR 2004.

[24] Y. Wei and L. Quan Region-Based Progressive Stereo Matching. CVPR 2004.

[25] M. Bleyer and M. Gelautz. A layered stereo algorithm using image segmentation and global visibility constraints. Submitted to ICIP 2004.

[26] Real-Time Stereo by using Dynamic Programming. Anonymous, submitted to CVPR 2004.

NO. OF
COPIES ORGANIZATION

* ADMINISTRATOR
DEFENSE TECHNICAL INFO CTR
ATTN DTIC OCA
8725 JOHN J KINGMAN RD STE 0944
FT BELVOIR VA 22060-6218
*pdf file only

1 DIRECTOR
US ARMY RSCH LABORATORY
ATTN IMNE ALC IMS MAIL & REC MGMT
2800 POWDER MILL RD
ADELPHI MD 20783-1197

1 DIRECTOR
US ARMY RSCH LABORATORY
ATTN AMSRD ARL CI OK TL
TECH LIB
2800 POWDER MILL RD
ADELPHI MD 20783-1197

2 DIRECTOR
US ARMY RSCH LABORATORY
ATTN AMSRD ARL SE SE P GILLESPIE
N NASRABADI
2800 POWDER MILL RD
ADELPHI MD 20783-1197

1 NATL INST OF STDS & TECHNOLOGY
ATTN DR M SHNEIER
BLDG 200
GAITHERSBURG MD 20899

2 UNIV OF MARYLAND
INST FOR ADV COMPUTER STUDIES
ATTN DR L DAVIS D DEMENTHON
COLLEGE PARK MD 20742-3251

ABERDEEN PROVING GROUND

1 DIRECTOR
US ARMY RSCH LABORATORY
ATTN AMSRD ARL CI OK (TECH LIB)
BLDG 4600

2 DIRECTOR
US ARMY RSCH LABORATORY
ATTN AMSRD ARL WM J SMITH
T ROSENBERGER
BLDG 4600

2 DIRECTOR
US ARMY RSCH LABORATORY
ATTN AMSRD ARL WM B W CIEPIELLA
BLDG 4600

NO. OF
COPIES ORGANIZATION

16 DIRECTOR
US ARMY RSCH LABORATORY
ATTN AMSRD ARL WM BF H EDGE
M BARANOSKI P FAZIO
M FIELDS G HAAS (4 CYS)
T HAUG W OBERLE (4 CYS)
R PEARSON R VON WAHLDE
S WILKERSON
BLDG 390

2 DIRECTOR
US ARMY RSCH LABORATORY
ATTN AMSRD ARL WM RP C SHOEMAKER
J BORNSTEIN
BLDG 1121