Stereo-Vision Based 3D Modeling of Space Structures

Stephen Se*[1], Piotr Jasiobedzki*, Richard Wildes**

*MDA, Space Missions, 9445 Airport Road, Brampton, Ontario, L6S 4J3, Canada
**Department of Computer Science, York University, 4700 Keele Street, Toronto, Ontario, M3J 1P3, Canada

ABSTRACT

Servicing satellites in space requires accurate and reliable 3D information. Such information can be used to create virtual models of space structures for inspection (geometry, surface flaws, and deployment of appendages), estimation of relative position and orientation of a target spacecraft during autonomous docking or satellite capture, replacement of serviceable modules, detection of unexpected objects and collisions. Existing space vision systems rely on assumptions to achieve the necessary performance and reliability. Future missions will require vision systems that can operate without visual targets and under less restricted operational conditions towards full autonomy.

Our vision system uses stereo cameras with a pattern projector and software to obtain reliable and accurate 3D information. It can process images from cameras mounted on a robotic arm end-effector on a space structure or a spacecraft. Image sequences can be acquired during relative camera motion, during fly-around of a spacecraft or motion of the arm. The system recovers the relative camera motion from the image sequence automatically without using spacecraft or arm telemetry. The 3D data computed can then be integrated to generate a calibrated photo-realistic 3D model of the space structure.

Feature-based and shape-based approaches for camera motion estimation have been developed and compared. Imaging effects on specular surfaces are introduced by space materials and illumination. With a pattern projector and redundant stereo cameras, the robustness and accuracy of stereo matching are improved as inconsistent 3D points are discarded. Experiments in our space vision facility show promising results and photo-realistic 3D models of scaled satellite replicas are created.

Keywords: Stereo vision, 3D modeling, Inspection, Specular surfaces, Space structures, Satellite servicing

1. INTRODUCTION

Servicing and inspection of space structures, satellites and spacecraft in space requires accurate and reliable three dimensional (3D) information at ranges from almost contact up to hundreds of meters. Such information may be used to create 3D models of space objects of interest for their inspection (measurement of their geometry, detection of surface flaws and verification of correct deployment of appendages), estimation of satellite pose during autonomous docking or capture, replacement of serviceable modules, detection of unexpected objects in the workspace, and detection of impending collisions.

3D imaging of space structures is difficult due to

- Imaging properties and shape of the space structures
- Illumination in space
- Uncertain relative camera location
- Relative motion of observed objects

Space structures such as satellites, space modules and spacecrafts are covered with thermal insulation and shields that provide some protection against micro-meteorites. Such materials often have partially specular characteristics or

[1] Author for correspondence – Stephen Se – Email: stephen.se@mdacorporation.com Telephone: 1 905 790 2800 x4270

Sensors and Systems for Space Applications, edited by Richard T. Howard, Robert D. Richards
Proc. of SPIE Vol. 6555, 65550E, (2007) · 0277-786X/07/$18 · doi: 10.1117/12.718427

homogenous appearance. The shapes of spacecraft and space structures are often complex. There are rarely any planar or simple surfaces, and there are protruding appendages and structures such as antennas, masts, solar arrays, thrusters and sensors. Intense and directional sun illumination creates hard shadows and high dynamic range scenes that often exceed the dynamic range of cameras. Earth albedo provides uniform and almost shadowless illumination.

When the servicer spacecraft is berthed with the satellite, then the servicer robotic arm can be used for servicing and inspection. However, as space robots are designed to be lightweight, have long reach, and operate with low power, the end-effector cannot be positioned accurately or without any oscillations. Therefore, using arm telemetry alone, for example, integration of multiple views from end-effector cameras and precise robotic tasks is not possible. Nevertheless, a vision system can be used, which can process images from end-effector cameras mounted on the robotic arm to provide accurate relative 3D measurements.

If the cameras are mounted on a free-floating or orbiting spacecraft, then only a coarse relative motion may be known. During inspection, the vision system must be able to compute accurate relative motion in order to allow the registration of multiple measurements. Similarly, relative motion or oscillation of the end-effector must be estimated to allow the registration of data from the end-effector mounted camera.

Inspection of deployed satellites requires creating “as-is” 3D surfaces models that can be visualized for human inspection including interactive measurements. The 3D models may be analyzed automatically to detect surface or shape flaws, partially deployed appendages, and to compare with prior 3D models. The created models must contain a sufficient amount of 3D detail and realistic surface appearance. The models may include multi-modal data obtained beyond the visual spectrum (IR, UV).

We have developed a stereo-vision based 3D modeling system “Vision3D”, which will be particularly applicable to observation of satellites, space station modules and structures, shuttle body covered with thermal tiles, etc. The data will be acquired using a lightweight mobile sensor that can be mounted on a robotic end-effector (operating at a camera to worksite stand-off distance of 0.5 to 3m) or a micro-satellite/servicer spacecraft (operating at a camera to worksite stand-off distance of 4 to 20m).

Our system integrates a novel sensor with algorithms and software that will process image sequences acquired under relative motion and compute reliable and accurate 3D data. The 3D sensor will consist of stereo cameras and an integrated pattern projector that will allow computing 3D data for diffuse, featureless and partially specular surfaces. This data will be converted into a photo-realistic 3D surface representation suitable for visualization, accurate measurement, flaw detection and comparison with prior models. The 3D data may also be used for model-based pose estimation, and detection of unexpected objects within the robotic workspace or impending collisions.

2. RELATED WORKS

2.1 Space Vision Systems

First generation space vision systems have relied on simplifying assumptions that allowed them to achieve the necessary performance and reliability using the processing hardware available. Such assumptions included using high contrast visual targets and elimination of all specular surfaces from the field of view, and restricting operational conditions to eclipse or certain viewing distances and angles. A human operator was often in control of the system initialization and parameter selection. Such assumptions allowed for the achievement of successful operations of several systems in technology demonstration missions: SVS [1], VGS, and ETS-VII [2]. However, these assumptions turned out to be too restrictive for many existing and future applications. Future missions will require vision systems that can operate with significantly relaxed assumptions: no visual targets or modifications to the space hardware, less restricted operations (any viewpoint, wide range of illuminations), and full vision system autonomy [3].

Our Vision3D system can automatically create 3D models of observed space structures with partially specular surfaces and does not rely on any visual targets for its operation.

2.2 3D Sensors

Among a wide range of existing and emerging 3D imaging technologies [4, 5], the followings are the most promising for space applications:

- Scanning rangefinders
- Flash LIDAR
- Structured light
- Multi-view stereo

Systems with mechanical scanners (scanning LIDAR/rangefinders) are generally too large and require too much power to be installed on lightweight robotic arms or micro-satellites. Furthermore, sequential data acquisition restricts operations in dynamic environments and requires precise motion estimation/compensation algorithms. Flash LIDAR overcomes the need for motion compensation, but the existing prototypes have low image resolution. Structured light systems do not require any scanning; however, the acquisition of dense data requires multiple scans, which restricts their use in dynamic environments. The reliable detection of encoded patterns may be difficult for specular space materials. Passive stereo systems require a natural surface texture, which is usually not present on spacecraft surfaces.

We have developed an active stereo vision system which addresses most of the requirements for a 3D imaging system for space. An integrated pattern projector projects a pattern on the scene providing an ample amount of texture for stereo matching. Using patterns and improved stereo matching algorithms helps to increase the 3D coverage and accuracy for specular surfaces. The discriminating factors compared to other reviewed systems are

- Low mass, size, power requirements, cost and no moving parts required for 3D data acquisition
- No visual targets on the observed object required
- Real time data acquisition, on-line or off-line processing depending on the computational platform
- Data acquisition during relative (sensor/object) motion
- High resolution 3D measurements and 2D images
- Photo-realistic and high resolution representation of the surface appearance

2.3 Dense Stereo

A point in a three-dimensional scene projects to different positions in two-dimensional images that are acquired from different viewpoints. Stereo vision is concerned with inverting this process to exploit the differing appearance of multiple views of a scene to recover the three-dimensional layout.

An outstanding challenge in the realization of automated stereo systems is the development of algorithms for calculating the correspondence between the input images accurately for surfaces with a wide range of imaging properties. Stereo methods can be classified into two classes: sparse feature-based methods and dense area-based methods. Sparse feature-based methods extract prominent features (e.g., intensity edges) from the individual images and then search over spatial position to match similar features across images [6, 7]. Such methods can yield reliable estimates at the feature points as they are well localized in the individual images and can be realized with efficient implementations, as the extracted features are sparse compared to the original images. A major limitation of such approaches is their inherent inability to provide estimates at non-feature points, i.e., the depth estimates are as sparse as the extracted features.

In contrast, dense area-based methods perform their matching directly on the input image data, e.g., in terms of spatially overlapping windows defined on the image intensity or a derived image selected so as to enhance local image patterns and structure (e.g., bandpass images). Area-based methods yield spatially dense estimates of depth at reasonable execution rates with recent advances in hardware capabilities [8]. A major limitation of such approaches is that the estimates are only reliable if the local image pattern is distinctive, e.g., poor estimates are common for homogenous surfaces. Another problem with such methods is that they tend to produce inaccurate results around object boundaries as the windows used to define matching regions may overlap foreground and background surfaces.

In addition, the explicit formulation of the correspondence process as a global optimization procedure allows non-local geometric context to help resolve matching ambiguities. Example approaches include those that formulate the problem in terms of dynamic programming [9], max-flow [10] and graph-cuts [11]. Such methods have provided some of the most accurate three-dimensional reconstructions from input imagery to date [12]. However, these results come at a higher computational expense and/or sensitivity to their parameter tunings in comparison to purely feature-based and area-based methods.

We have developed a novel 3D sensor and dense stereo matching algorithm that improves the 3D coverage and accuracy for objects with partially specular and featureless surfaces.

2.4 Motion Estimation

There are various approaches for motion estimation from images. A feature-based approach is used in [15], where SIFT features (Scale Invariant Feature Transform) [14] are extracted from stereo images and these features are matched over frames to recover the camera ego-motion. The accuracy of the motion estimation depends on the number of features and their distribution in the scene.

The shape-based approach can alternatively be employed for motion estimation by aligning the latest 3D data set with the previous one or with the overall cumulative 3D data set. Iterative Closest Points (ICP) [16] is a common technique for aligning two sets of 3D data by minimizing the geometrical distance iteratively. This algorithm consists of multiple steps which are iterated until termination conditions (reaching a minimum RMS or maximum number of allowed iterations) are met. The following key steps are executed iteratively:

- Computation of the closest points between two sets of 3D points
- Computation of geometrical registration between the data points
- Application of geometrical registration to the data

Acceleration techniques that rely on partitioning the data as k-d trees and structuring the computation [17, 18] have been proposed. A review of recent research in the ICP class of algorithms can be found in [19]. At MDA, we have previously developed fast ICP algorithms for model-based pose estimation [20, 21].

In Vision3D system we employ the ICP algorithm to align 3D points obtained from dense stereo to build the 3D model. We have obtained good results from testing with scaled satellite replicas such as Radarsat and Hubble Space Telescope.

3. SYSTEM ARCHITECTURE

3.1 Background

Object and environmental modeling can be regarded as two approaches of 3D modeling, with an outside-looking-in approach more suitable for object modeling and an inside-looking-out approach more suitable for environmental modeling.

The Vision3D system uses components of the MDA's instant Scene Modeler (iSM), a 3D modeling system developed for environmental modeling [13]. iSM automatically processes sequences of stereo images obtained from a calibrated stereo camera and creates calibrated photo-realistic 3D models. iSM uses a feature-based approach which detects tie points in images for ego-motion estimation [15]. The system typically detects hundreds of SIFT tie points in images, that allows for good motion estimation when the cameras observe environments. As man-made objects such as space structures usually lack natural features and objects typically only occupy a portion of the camera field of view, the SIFT-based estimated motion may not be accurate and hence this approach is more suitable for environmental modeling, rather than object modeling.

3D data computed for each stereo pair is transformed into the initial coordinate frame using the estimated motion. The transformed data is integrated and used to create a 3D surface, which is enhanced by texture mapping using selected camera images. This process is fully automatic. The motion estimation may optionally use data provided from external devices to improve the estimates obtained from processing images, e.g., orientation sensors, manipulator telemetry.

Using all the 3D points obtained from the stereo processing is not efficient as there are too many redundant measurements, and the data may contain noise and missing regions (due to incorrect matches or lack of texture). Representing 3D data as a triangular mesh reduces the amount of data when multiple sets of 3D points are combined. Furthermore, creating surface meshes fills up small holes and eliminates outliers, resulting in smoother and more realistic reconstructions.

To generate triangular meshes as 3D models, we employ a voxel-based method [22], which accumulates 3D points with their associated normals. It creates a mesh using all the 3D points, fills up holes and works well for data with significant overlap. The 3D data is accumulated into voxels at each frame. Outliers are filtered out using their local orientation and by selecting the threshold of range measurements required per voxel for a valid mesh vertex. It takes a few seconds to construct the triangular mesh at the end, which is dependent on the data size and the voxel resolution.

3.2 System Overview

The system architecture for the Vision3D system is shown in Figure 1. The images are processed automatically to generate a photo-realistic 3D model for visualization and post-processing.

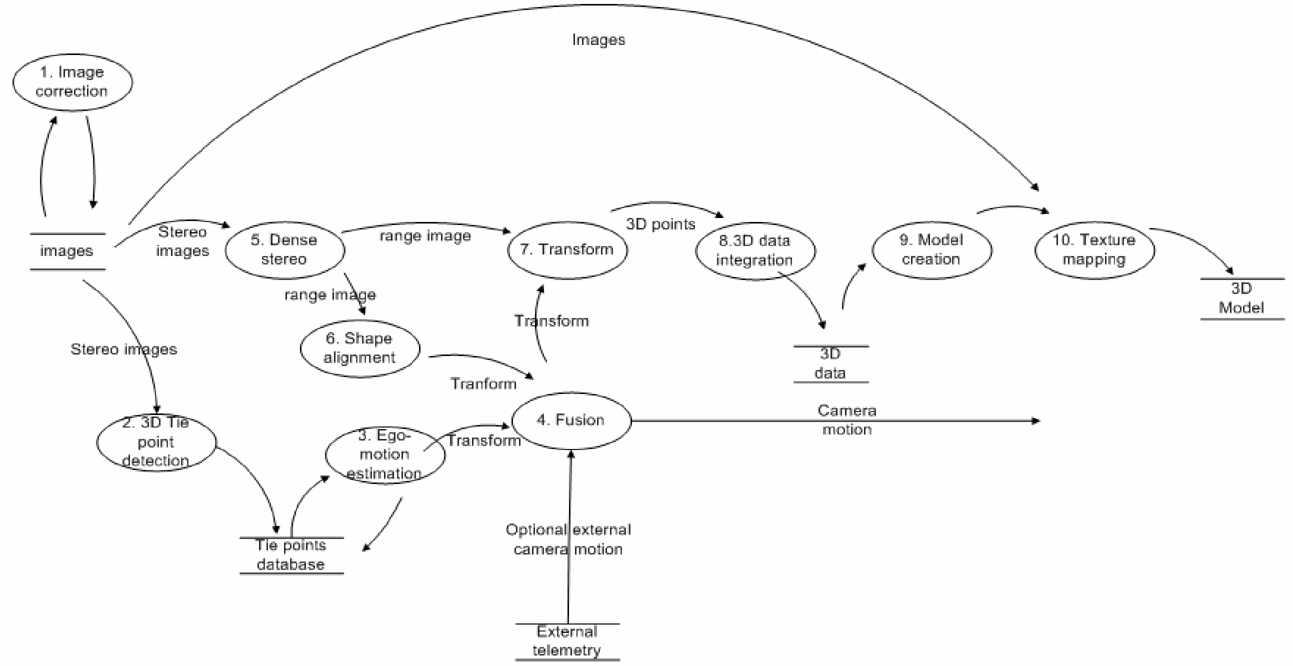

Figure 1 Vision3D system architecture

To capture the image sequence, the Vision3D system sends flash request to the pattern projector, which is synchronized with the stereo cameras. The stereo images are used for both motion estimation and dense stereo processing to generate 3D range data. Rather than using 3D tie points for motion estimation, shape alignment based on the 3D range data is performed. External telemetry may be provided if the relative motion between the camera and observed object is available. The range data is transformed into the initial camera coordinates frame according to the estimated camera motion. This is followed by the voxel-based triangular mesh creation and texture mapping, as in iSM.

The key advancements from iSM to Vision3D are as follows and they will be described in detail in the next few sections.

- 3D sensor to improve 3D coverage and accuracy, particularly for an outside-looking-in object modeling.
- Dense stereo algorithm to handle partially specular surfaces
- Motion estimation by shape alignment for object modeling

3.3 3D Sensor

With a passive stereo camera, poor 3D coverage is obtained for featureless surfaces. A pattern flash projector that projects a random pattern onto a scene allows standard stereo algorithms to compute dense 3D representation even for featureless scenes. The flash is synchronized with the cameras allowing images to be captured with or without the pattern.

Stereo images were captured for the flight hardware of the Special Purpose Dexterous Manipulator (SPDM) with and without the pattern projector. The SPDM has homogeneous regions as well as specular regions. The dense stereo matching results for both cases are shown in Figure 2. The grey regions correspond to pixels where the stereo algorithm was able to compute reliable data (brighter intensities represent features closer to the camera and black where no data was computed). Using the passive stereo system and local dense stereo algorithms results in a sparse representation of the scene with data computed only for edges and textured regions (e.g., CSA logo). Using the pattern projector allows recovering 3D data for featureless regions. Some data was lost in the highly specular regions due to saturation and reflections.

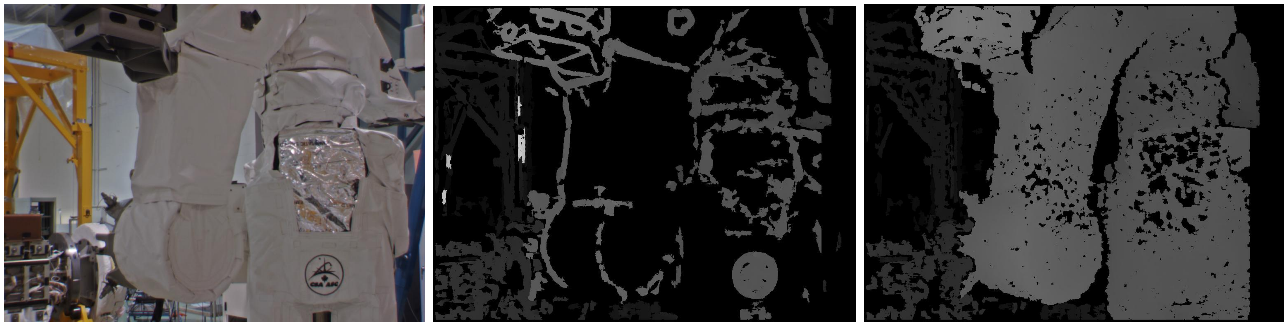

Figure 2 Special Purpose Dexterous Manipulator (SPDM) on a test stand (left), dense 3D data computed without pattern projector (middle) and with pattern projector (right). Much denser 3D points are obtained with the use of the projector.

To develop a reconstruction method that is robust to specular reflection, it is useful to consider a model of surface reflection. For example, the following model of surface reflectance, E, is widely used in the computer graphics community to capture the appearance of surfaces that have both specular and matte components [23].

$$E = \rho\lambda\left[\alpha(\mathbf{n}\cdot\mathbf{s}) + (1-\alpha)(\mathbf{n}\cdot\mathbf{h})^k\right]$$

$\rho \equiv$ surface albedo

$\lambda \equiv$ source intensity

$\alpha \equiv$ matte vs. specular weighting

$\mathbf{n} \equiv$ unit surface normal

$\mathbf{s} \equiv$ unit source direction

$\mathbf{h} \equiv$ unit bisectrix of source and view directions

$k \equiv$ specular exponent

In this model the overall surface reflectance is given as a convex combination of matte, **n.s**, and specular, **n.h**, components, which capture the underlying surface appearance and reflections of the light source, respectively. Notice that the matte component obeys Lambert's Law, while the specular component obeys mirror reflection. The matte component does not change as a function of view direction; whereas, the specular component does change (**h** is a function of view direction and **s**). These observations suggest that matches between features across multiple views of a scene that are geometrically consistent will arise from the matte component of reflectance, while those that are inconsistent (as they change with viewpoint) will arise from the specular component.

In order to handle specular regions, we have investigated different configurations of stereo cameras and pattern projectors. Different optical arrangements have been tested by acquiring images of a test object with mixed specular and diffuse surfaces to select the most appropriate configuration.

1. 1 stereo camera with 1 static projector

   This improves the 3D coverage compared with a stereo camera with no projector, but it cannot handle specular regions on the object.

2. 1 stereo camera with 1 dynamic projector

   Specular reflections appear at different points on a surface depending on the viewpoint. Dynamically altering the projected pattern does not help, as the pattern is from the same direction and neither the source direction nor the view direction changes. In order for specularities to move across the surface, either the source direction or the source/view bisectix needs to vary.

3. 1 stereo camera with multiple static projectors

   By varying the source direction in this case, the specularities can be moved across the surface. However, this affects the matte component as well as the specular component. It is better to vary the source/view bisectix rather than the source direction, so that the underlying matte surface reflectance remains unaltered for establishing correct correspondences. Therefore, rather than varying the source location, multiple view points are needed.

4. 2 stereo cameras with 1 static projector

   In this case, the view points vary and hence the source/view bisectix changes and specularities move across the surface. Two stereo cameras are used and hence four images are captured from slightly different views. This is the chosen configuration for Vision3D. We use two Bumblebee stereo cameras from Point Grey Research [24] and a custom-built pattern projector for our prototype camera head, as shown in Figure 3.

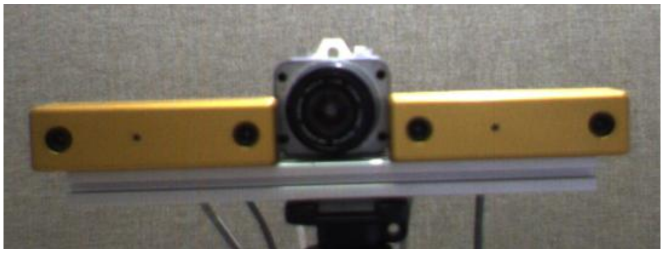

Figure 3 Prototype camera head for 2 stereo cameras with 1 projector configuration

3.4 Dense Stereo from Multiple Views

The standard stereo approach uses a pair of stereo images and implicitly assumes the Lambertian (non-specular) reflection model. The light reflections off specular surfaces (non-Lambertian reflection) create artifacts or data loss. In Vision3D, we have developed and characterized a new algorithm that processes multiple stereo images of the same objects to reduce the effects of specularities.

In general, three-dimensional scene structure can be recovered from multiple two-dimensional views of a scene through a three-step procedure. First, the cameras are calibrated so that their internal geometries and relative geometries are known. Second, matches between points in the images are established that arise from the same points in the world. This produces the disparity image which indicates the pixel difference of each 3D point seen in the left and right images. Third, rays are backprojected from the matched points in the images to obtain their intersections as the corresponding 3D locations for the image points. The stated paradigm fails in the presence of specular reflections in the scene. In such cases, the apparent features that appear in the images correspond not to the surfaces of interest, but rather to virtual images of the scene illuminants. Attempts to apply the standard multiple view approach to reconstruction fail to capture the geometry of the underlying surface.

Based on the surface reflection model above, an algorithm has been developed for analyzing a set of point matches that have been established between multiple views of a scene so that geometrically inconsistent matches are rejected, while consistent matches are fused to yield a single set of matches that are devoid of corruption due to specular reflections. Once this fusion has been accomplished, it is once again possible to reconstruct the 3D scene geometry through back projection of matched features.

As two stereo cameras are used, we have four images. As the individual cameras are calibrated beforehand, the four images can be rectified together. Three pair-wise dense stereo disparity maps between neighboring cameras are recovered and then all the information is combined via a consensus analysis.

The algorithm works as follows:

- Input: A set of 3 disparity maps and co-registered confidence maps resulting from matching feature points between adjacent rectified images acquired from 4 camera images of the same scene; also, a match consistency threshold. Let d12, d23 and d34 be disparity maps relating images 1 and 2, images 2 and 3, images 3 and 4, respectively.

- Warp all disparity maps to image 4 coordinates. Disparity maps are used to define the required warping. For example, to warp d12 to image 4 coordinates, we first (backward) warp d12 according to d23 and subsequently according to d34.
- For each spatial location in image 4 coordinates, consider all three available disparity estimates. Find the largest subset of the three local estimates that lie within the supplied match consistency distance of one another. Let the selected disparities be called the local consistency set.
- Output a single disparity map from the weighted combination of the local consistency set, using the co-registered confidence maps to provide the weighting coefficients.

For further robustness, disparity maps can be warped both forward and backward into a common coordinate frame to ameliorate difficulties encountered in warpings that are themselves defined via locally specular matches. Moreover, a maximum of six pair-wise dense stereo disparity maps can be computed from the four images. Furthermore, additional images acquired from different positions will allow for more data points to be considered in the consistency analysis.

3.5 Motion Estimation

As the SIFT feature-based approach does not work well with object modeling, 3D data points from the dense stereo computation are used for shape alignment to estimate the camera motion. This occurs for images with a flash pattern where sufficiently many 3D points are obtained. A modified ICP algorithm [16] is used to align the current 3D points with the previous 3D points. For normal images without flash, camera pose will be interpolated using the camera pose before and after at the fusion step. The normal images are required as the texture of the 3D model. Therefore, an image sequence consists of interleaving images with flash and images without flash. The pose obtained from the previous frame is used as the initial estimate for ICP this frame, for slow relative motion between frames.

Two ICP modifications have been implemented to improve the performance in terms of accuracy and speed. The computation of the closest points depends on the distance criterion. Previously, it was a fixed number, for instance, points within 200mm were considered. However, non-overlapping points may shift the points incorrectly. Therefore, our ICP distance criterion varies at each iteration depending on the RMS error of the previous iteration.

In the beginning, the correct matches could still be far away and hence a relatively large criterion is used. But as it iterates, the RMS error gets smaller and we shrink the distance criterion according to the RMS error of the last iteration. Therefore, only the closer points will be considered and this will avoid non-overlapping points from affecting the alignment. This has resulted in an improvement on alignment accuracy.

Moreover, there are typically hundreds of thousands of 3D points at each frame. Performing ICP between 2 large sets of points takes a large amount of computation, as there are many points to be matched. Therefore, we vary the number of points by sub-sampling the points at each iteration. In the early iterations, using a small percentage of the points is sufficient to bring the 2 sets of points into a rough alignment. In the latter iterations, all the points are used to obtain the fine alignment.

As the RMS error should be monotonically decreasing for a typical convergence case, we vary the number of points based on the RMS error. Starting with 1000 points, when the RMS error decreases only very slightly (less than 0.05%), the number of points to be considered is doubled, as not much alignment improvement can be obtained otherwise. Typically, less than 10% of the total numbers of points are used in most of the iterations, and a large proportion of the points will only be used in the last few iterations. In this manner, accurate alignment can still be obtained while keeping the computational requirement low.

4. EXPERIMENTAL RESULTS

3D coverage and accuracy of the dense stereo matching have been tested using the 3D sensor in conjunction with the dense stereo software for different configurations and illuminations. These tests have been performed using various test objects with specularities. Motion estimation has been tested for selected objects at MDA’s Space Vision test facility to compare the feature-based and shape-based approaches. The facility includes two Fanuc industrial robots that are used to effect relative motion between scale models of space objects or spacecrafts and the Vision3D active vision system hardware. One of the Fanuc robots holds the camera head while the other holds the test object. The robot holding the camera head was commanded to follow pre-defined trajectories to acquire images. As the actual pose of the camera is known from robot telemetry, it allows quantitative comparison of the algorithms. 3D model accuracy has been tested as part of the complete system tests.

4.1 Dense Stereo

The following 3 configurations were compared:

- 1 Bumblebee camera without projector (1BBnoP)
- 1 Bumblebee camera with projector (1BBwithP)
- 2 Bumblebee cameras and pattern projector (2BBwithP)

Several objects were tested and the results were consistent for both spot and ambient illumination. Typically, the 1BBnoP disparity images contain some wrong disparity values. With the pattern projector, the 1BBwithP disparity images contain fewer wrong disparity values as well as higher coverage thanks to the projected pattern. The 2BBwithP disparity images further removes the wrong disparity values but at the expense of slightly lower coverage as it seeks for consensus across all 4 images. The percentage of the coverage and the number of mismatches (in parenthesis) are tabulated in Table 1.

Table 1 Disparity coverage and number of mismatches of different objects for different configurations and illumination

Illumination	Configuration	Connector Worksite	Plane	Docking Interface	Radarsat Mock-up
Spot lighting	1BBnoP	19.5% (1760)	10.8% (377)	39.7% (3301)	18.6% (1497)
	1BBwithP	21.4% (999)	20.7% (53)	45.2% (691)	19.9% (426)
	2BBwithP	18.3% (0)	19.7% (0)	39.8% (0)	16.5% (59)
Ambient lighting	1BBnoP	23.6% (0)	22.0% (2404)	42.4% (2319)	20.0% (1372)
	1BBwithP	30.2% (261)	21.1% (263)	46.8% (228)	23.3% (209)
	2BBwithP	26.2% (0)	20.0% (0)	42.6% (0)	19.3% (0)

On a Pentium IV 3.4 GHz computer, the processing time for 2BBwithP (around 12 seconds) is much longer than the one Bumblebee camera cases (around 1.3 second) at 1024x768 image resolution. This is expected since there are 4 cameras available for the 2 Bumblebee cameras case, and pair-wise dense stereo matching is carried out three times, followed by a consensus analysis to combine the outputs from the pair-wise dense stereo matching. As more camera information is utilized, 3D data of better quality can be obtained, but at the expense of increased computational requirements.

4.2 Motion Estimation

The original pose estimation algorithm used for instant Scene Modeler (iSM) is based on matching SIFT features. Vision3D uses the ICP algorithm instead to compute the relative pose. The motion estimation accuracy is compared between the SIFT-based approach and the ICP-based approach. Ground truth information is available from the Fanuc robot telemetry and the robot is commanded to translate/rotate a known amount between frames.

Tests have been carried out using scale models of the Radarsat I and Hubble Space Telescope spacecrafts against two different backgrounds for various trajectories. The backgrounds include a clear background and a cluttered background. The results shown in Figure 4 are for Radarsat with a rotate & translate trajectory where it rotates 1 degree and translates 30mm between frames.

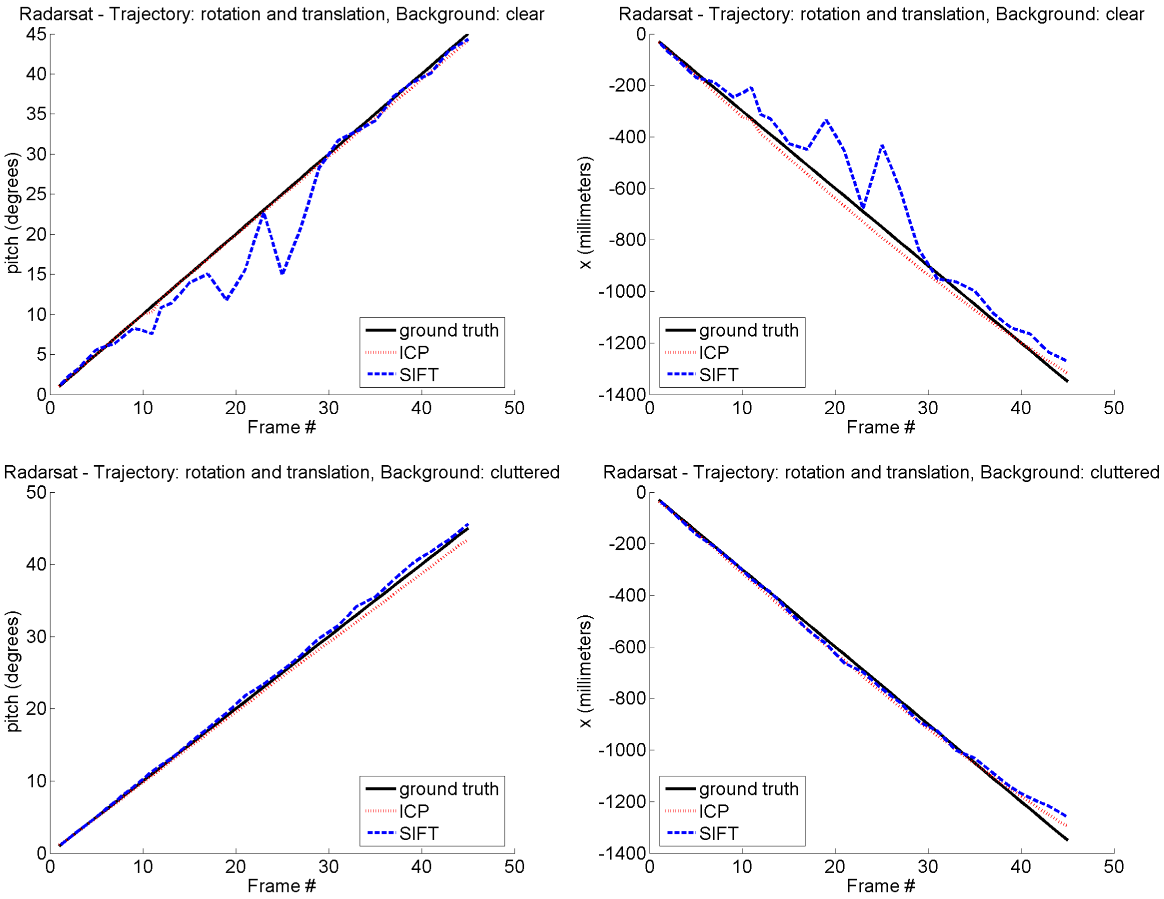

Figure 4 Motion estimation accuracy for Radarsat mock-up for the rotate & translate trajectory: clear background (top) and cluttered background (bottom). ICP results follow the ground truth better than the SIFT results for clear background.

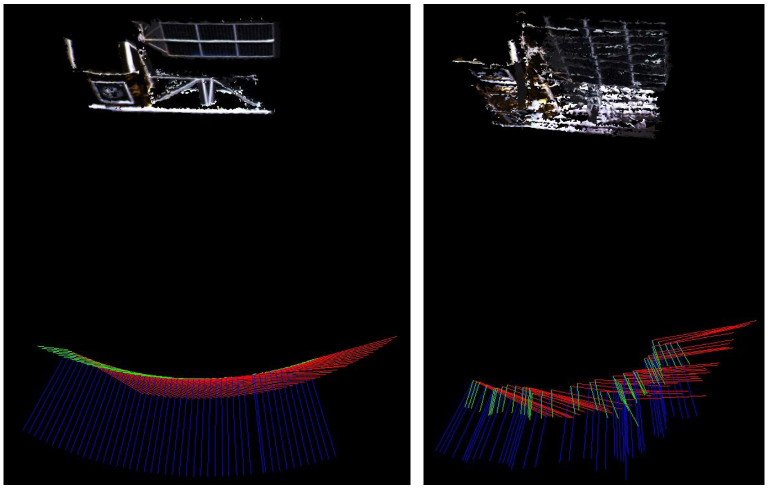

Figure 5 Screenshots showing the recovered trajectory and reconstructed Radarsat 3D model for clear background: ICP (left) and SIFT (right) approaches. A cleaner 3D model and a smoother trajectory are obtained with the ICP approach.

With a clear background, ICP results are close to the ground truth and it performs better than the SIFT approach. The recovered trajectories together with the reconstructed 3D model are shown in Figure 5. We can see that the recovered trajectory for the ICP case is smooth and the 3D model is clean. On the other hand, for the SIFT approach, the recovered trajectory is very rough and the 3D model is very noisy as well, due to the error in the camera pose estimation.

With the cluttered background, both ICP and SIFT approaches perform well. The improvement with the SIFT approach for the cluttered background is due to the extra SIFT features found in the background. As there are more features and they are more widely spread in 3D, the pose estimation accuracy is improved. However, in typical space scenarios, there will not be much background clutter and hence the ICP approach will perform better. It can be seen that the error grows towards the end of the sequence, due to error accumulation over the frames.

4.3 System Test

With the cameras at around 4 meters from the test objects, 3D models of the Radarsat are created and distance measurements are performed on the model to compare with the ground truth. The measurements are typically very close to the ground truth. Most of the measurements are within 5% of the ground truth. The errors include the stereo uncertainty, motion estimation error as well as the measurement error, i.e., where the point is selected on the 3D model.

The processing time required to generate 3D models depends on the object scene, the number of frames in the sequence as well as the camera trajectory. For the Rotate & Translate trajectory, a total of 47 frames were captured and the total processing time on a Pentium IV 3.4 GHz computer is 370 seconds, i.e., around 8 seconds per frame.

4.4 Loop Closure Trajectory

In this test, the Radarsat mock-up was rotated 360 degrees while the cameras remain stationary. Figure 6 shows some images from this trajectory sequence. As the camera axis is not aligned with the Radarsat axis, we do not have ground truth information for each frame, except that we know that the initial pose and final pose at the end of the sequence must be equal.

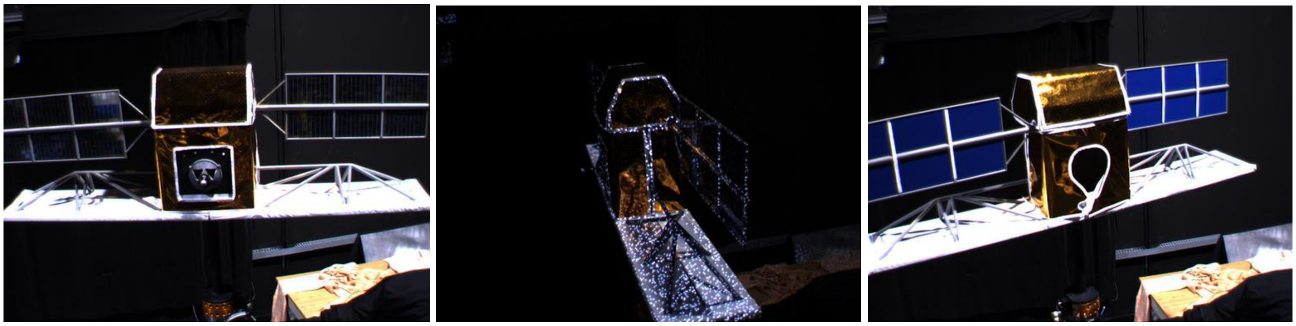

Figure 6 Images from Radarsat mock-up 360 degree trajectory sequence

The image sequence was processed in 3 modes:

- 1 Bumblebee stereo camera with the ICP approach (1BB + ICP)
- 2 Bumblebee stereo cameras with the ICP approach (2BB + ICP)
- 1 Bumblebee stereo camera with the SIFT approach (1BB + SIFT)

Figure 7 shows the recovered trajectories and 3D models for the different modes. The errors for the SIFT approach are so large that the recovered trajectory and the reconstructed 3D model are completely wrong. The 1 Bumblebee camera with the ICP result is much better, as the trajectory looks more or less correct, i.e., the camera goes around the object. Due to error accumulation over the sequence, the trajectory closure is not perfect and the 3D models also show some artifacts, as the same 3D points observed in the beginning and at the end will appear at slightly different locations. Compared to the 1 Bumblebee camera mode, the 2 Bumblebee camera mode has a smaller discontinuity in the trajectory and produces a cleaner 3D model.

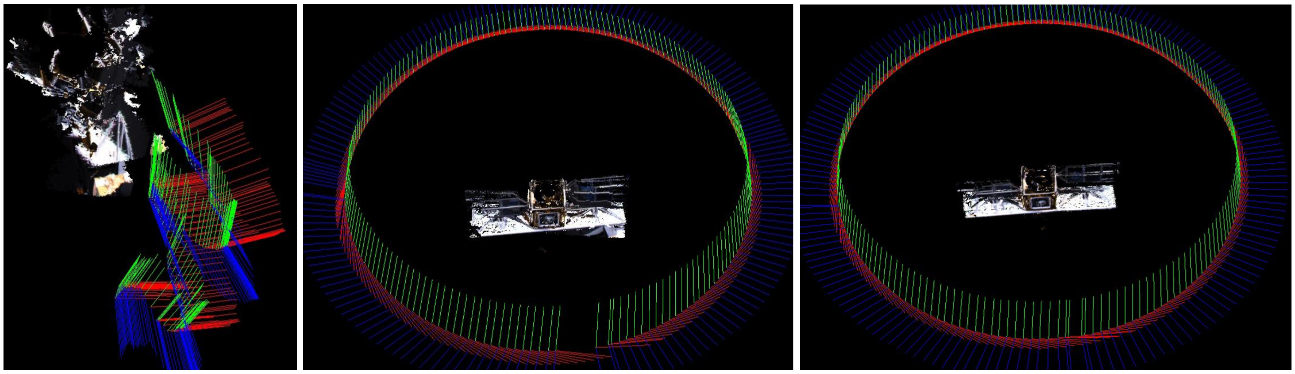

Figure 7 Recovered trajectories and 3D models for the Radarsat 360 degree sequence: 1 Bumblebee with SIFT (left), 1 Bumblebee with ICP (middle) and 2 Bumblebee cameras with ICP (right). The 2 Bumblebee cameras with ICP mode gives the best 3D model and trajectory, while the 1 Bumblebee with SIFT mode fails to generate the correct 3D model.

5. CONCLUSIONS

In this paper, we have presented Vision3D, a 3D modeling system for space structures. Vision3D's outside-looking-in approach for object modeling is complementary to iSM's [13] inside-looking-out approach for environmental modeling. The Vision3D system creates photo-realistic 3D calibrated models of space structures automatically within minutes using stereo cameras and a pattern projector. The cameras can move around the object or the object can rotate itself.

In order to handle specular regions, images are captured from different viewpoints and hence, the configuration of two stereo cameras with a static pattern projector has been chosen. The pattern projector improves the coverage and quality of the 3D data by projecting virtual texture onto the scene. The configuration of two stereo cameras with the pattern projector improves the quality of the 3D data by discarding inconsistent disparities due to specular surfaces. The processing time required is longer because four images are captured and three pair-wise dense stereo matching is carried out followed by consensus analysis.

For motion estimation, the SIFT approach works well in the iSM for environmental modeling, but the ICP approach in general performs better for object modeling. In particular, when the object is rotating while the camera is stationary, ICP can estimate the camera pose reasonably well, while SIFT cannot. The accuracy of the 3D models depends on the stereo uncertainty and motion estimation error throughout the image sequence. The measurements on reconstructed 3D models of test objects are within 5% to the ground truth distances on the objects in most cases.

The experimental results show that the use of a pattern projector and multiple stereo cameras improves the coverage and quality of the 3D data for space structures even with partially specular surfaces. With the ICP shape alignment, 3D data from the stereo image sequence can be integrated together to create accurate, photo-realistic 3D models of the test objects, including satellite mock-ups.

ACKNOWLEDGEMENTS

This project was funded in part by Canadian Space Agency's Space Technologies Development Program (STDP).

REFERENCES

1. J. McCarthy, Space vision system, International Symposium on Robotics (ISR), Birmingham, UK, 1998.
2. M. Oda, K. Kine, and F. Yamagata, ETS-VII – a rendezvous docking and space robot technology experiment satellite, International Astronautical Congress, pages 1-9, Norway, Oct 1995.
3. S. Hollander, Autonomous space robotics: enabling technologies for advanced space platforms, AIAA Space 2000 Conference, California, September 2000.
4. P. Besl, Active optical range imaging sensors, Machine Vision and Application, 1:127-152, 1988.

5. M. Hebert, Active and passive range sensing for robotics, IEEE International Conference on Robotics and Automation, 2000.
6. W.E.L. Grimson, Computational experiments with a feature based stereo algorithm, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 7, no. 11, pages 17-34, 1985.
7. S. Pollard, J. Mayhew, and J. Frisby, PMF: A stereo correspondence algorithm using the disparity gradient limit, Perception 14:449-470, 1985.
8. H. Hirschmuller, Improvements in real-time correlation-based stereo vision, IEEE Workshop on Stereo and Multibaseline Vision, 2001.
9. Y. Ohta and T. Kanade, Stereo by intra- and inter-scanline search using dynamic programming, IEEE Transactions on Pattern Analysis and Machine Intelligence 7(2):139-154, 1985.
10. S. Roy, Stereo without epipolar lines: a maximum flow formulation, International Journal of Computer Vision, 34(2/3):147-161, 1999.
11. V. Kolmogorov and R. Zabih, Computing visual correspondences with occlusions using graph cuts, IEEE International Conference on Computer Vision (ICCV), 2001.
12. D. Scharstein and R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, International Journal of Computer Vision, vol. 47, 2002.
13. S. Se and P. Jasiobedzki, Photo-realistic 3D model reconstruction, IEEE International Conference on Robotics and Automation (ICRA), pages 3076-3082, Orlando, Florida, May 2006.
14. D.G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision, vol. 60, no. 2, pages 91-110, 2004.
15. S. Se, D. Lowe, and J. Little, Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks, International Journal of Robotics Research, vol. 21, no. 8, pages 735–758, August 2002.
16. P. Besl and N. McKay, A Method for Registration of 3-D Shapes, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 2, pages 239-256, 1992.
17. D. Simon, M. Hebert, and T. Kanade, Real-time 3D Pose Estimation Using a High Speed Range Sensor, IEEE International Conference on Robotics and Automation, pages 2235-2241, 1996.
18. Z. Zhang, Iterative Point Matching for Registration of Free-Form Curves and Surfaces, International Journal on Computer Vision, vol. 13, no. 2, pages 119-152, 1994.
19. S. Rusinkiewicz and M. Levoy, Efficient Variants of the ICP Algorithm, International Conference on 3D Digital Imaging and Modeling (3DIM), 2001.
20. M. Abraham, P. Jasiobedzki, and M. Umasuthan, Robust 3D Vision for Autonomous Space Robotic Operations, 6th International Symposium of Artificial Intelligence and Robotics in Space (iSAIRAS), June 2001.
21. P. Jasiobedzki, J. Talbot, and M. Abraham, Fast 3D pose estimation for on-orbit robotics, International Symposium on Robotics (ISR), Montreal, Canada, May 2000.
22. G. Roth and E. Wibowo, An efficient volumetric method for building closed triangular meshes from 3-D image and point data, Graphics Interface (GI), pages 173–180, Kelowna, B.C., Canada, 1997.
23. B.T. Phong, Illumination for computer generated pictures, Communications of the ACM, 18(6):311-317, June 1975.
24. Point Grey Research, http://www.ptgrey.com