Cagatay Basdogan*
College of Engineering
Koc University
Istanbul, Turkey, 34450

From 2D Images to 3D Tangible Models: Autostereoscopic and Haptic Visualization of Martian Rocks in Virtual Environments

Abstract

A planetary rover acquires a large collection of images while exploring its surrounding environment. For example, 2D stereo images of the Martian surface captured by the lander and the Sojourner rover during the Mars Pathfinder mission in 1997 were transmitted to Earth for scientific analysis and navigation planning. Due to the limited memory and computational power of the Sojourner rover, most of the images were captured by the lander and then transmitted to Earth directly for processing. If these images were merged together at the rover site to reconstruct a 3D representation of the rover's environment using its on-board resources, more information could potentially be transmitted to Earth in a compact manner. However, construction of a 3D model from multiple views is a highly challenging task to accomplish even for the new generation rovers (Spirit and Opportunity) running on the Mars surface at the time this article was written. Moreover, low transmission rates and communication intervals between Earth and Mars make the transmission of any data more difficult. We propose a robust and computationally efficient method for progressive transmission of multi-resolution 3D models of Martian rocks and soil reconstructed from a series of stereo images. For visualization of these models on Earth, we have developed a new multimodal visualization setup that integrates vision and touch. Our scheme for 3D reconstruction of Martian rocks from 2D images for visualization on Earth involves four main steps: a) acquisition of scans: depth maps are generated from stereo images, b) integration of scans: the scans are correctly positioned and oriented with respect to each other and fused to construct a 3D volumetric representation of the rocks using an octree, c) transmission: the volumetric data is encoded and progressively transmitted to Earth, d) visualization: a surface model is reconstructed from the transmitted data on Earth and displayed to a user through a new autostereoscopic visualization table and a haptic device for providing touch feedback. To test the practical utility of our approach, we first captured a sequence of stereo images of a rock surface from various viewpoints in JPL MarsYard using a mobile cart and then performed a series of 3D reconstruction experiments. In this paper, we discuss the steps of our reconstruction process, our multimodal visualization system, and the tradeoffs that have to be made to transmit multiresolution 3D models to Earth in an efficient manner under the constraints of limited computational resources, low transmission rate, and communication interval between Earth and Mars.

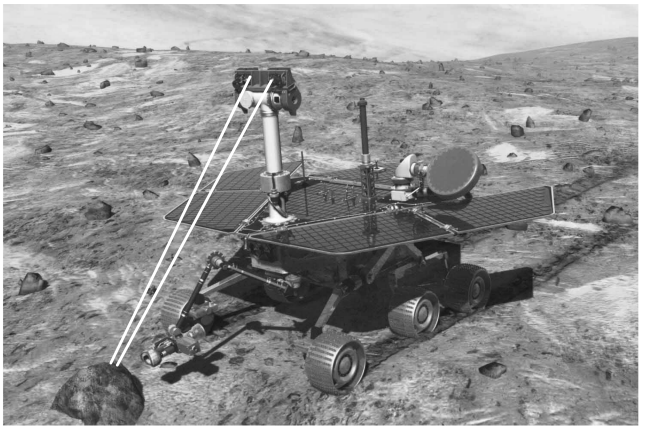
Presence, *Vol. 16, No. 1, February 2007,* 1–15

*Correspondence to cbasdogan@ku.edu.tr.

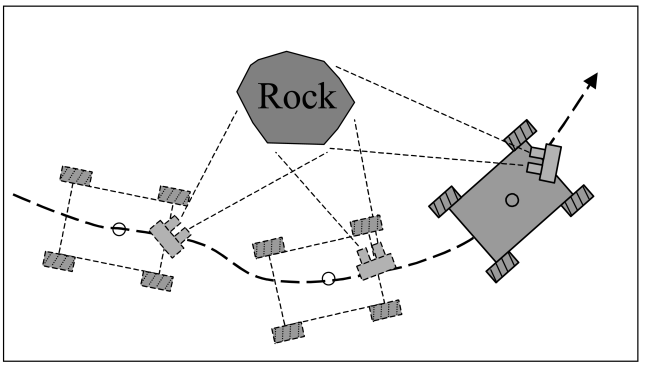
1 Introduction

During the Pathfinder mission to Mars, the stereo images captured by the lander and Sojourner rover were transmitted to Earth for scientific analysis and mission planning. Due to the limited memory and computational power of the processors of the Sojourner rover (the Sojourner rover was equipped with a 2 MHz CPU and 768 kB of memory), it was not possible to transmit many images to Earth or to process the captured images at the remote site. However, new generation rovers have increasingly more memory and computational power (e.g., Spirit and Opportunity rovers have a CPU power of 20 MHz and 128 MB RAM). They will be challenged to construct 3D representations of their environment from stereo images to transmit more planetary data to Earth in a compact manner and to make autonomous decisions for navigation. To respond to this challenge, efficient reconstruction algorithms that will work with the limited computational resources of the rover must be developed. In addition, progressive transmission protocols are necessary for the efficient transmission of the reconstructed models from the rover site to Earth. Progressive transmission of data is important because the communication interval and the data transmission rates between a planetary rover and Earth are very limited. (The data transmission rate was approximately 40 bits/s from Mars to Earth during the Pathfinder mission. The transmission rate for Spirit and Opportunity rovers has not been announced officially yet, but is reported to vary between 10 kbits/s and 120 kbits/s.)

We developed an integrated approach to construct multiresolution 3D models of Martian rocks from multiple range images for visualization on Earth. Vergauwen, Pollefeys, and Van Gool (2003) describes a method for 3D reconstruction of Martian rocks using stereo images. We further expand this concept and propose an end-to-end solution for progressive transmission and haptic visualization of 3D models of Martian rocks in virtual environments for scientific analysis. In our approach, 2D views of the rock and soil surfaces are obtained through stereo cameras mounted on a planetary rover (see Figure 1 a). In order to construct a detailed 3D representation of a Martian rock, the planetary rover captures stereo images of the rock surface from a number of different viewpoints (Figure 1b). Then, depth maps, also known as range maps, are constructed from stereo images at the remote site using a simple triangulation method. Each point in range maps contains the 3D coordinates of the rock surface relative to the position and orientation of the rover. We then implement four efficient steps (a, b, c, d) to construct 3D models of the rock from these range maps for visualization on Earth (see Figure 2):



(a)



(b)

Figure 1. *a) The Mars Exploration Rover Project at JPL launched two identical rovers (Spirit and Opportunity) in mid-2003 for NASA to explore the Martian surface. Following the landing, scientists commanded the rovers to go to rock and soil targets of interest for capturing their images and evaluating their composition and texture. b) Typically, a rover cannot make a full tour around a rock to capture all its views due to the mission constraints.*

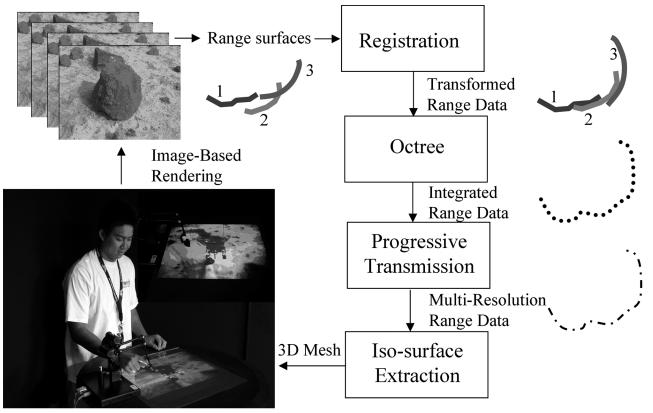


Figure 2. *The steps of our 3D reconstruction process: a) registration of range maps, b) octree-based data representation, c) progressive transmission, and d) autostereoscopic and haptic visualization of planetary data in virtual environments.*

a) Registration. The first step is to register all these individual range maps so that the range points are in the same coordinate frame. Since the position and orientation of the rover change relative to the rock as the rover moves, successive range maps cannot be combined without registration. This process can be done using the range data itself by aligning the overlapping regions such that the total error between the range points is minimized.

b) Integration. The second step is to construct a geometric representation of the rock from the properly aligned range maps. This step is also called range integration since the goal is to fuse all the registered maps into a global representation. We used an octree structure to integrate the registered range points into a single volumetric representation.

c) Transmission. The integrated data is transmitted to Earth progressively in increasing resolutions. Progressive data transmission enables the mission scientists to visualize the low-resolution representations of the rock and Martian soil before a full resolution model becomes available. Scientists can then decide to continue or stop the transmission based on what they have seen so far.

d) Visualization. The final step involves the reconstruction of multiresolution 3D surface models of the rock from the transmitted data on Earth and autostereoscopic and haptic visualization of the model in a virtual environment for scientific analysis.

At JPL, methods have been developed for on-board science processing such as image-based prioritization of Martian rocks (i.e., deciding which rock to investigate; Manduchi, Castano, & Cohen, 2001) and progressive transmission of these images to Earth for more detailed scientific analysis (Kiely, Manduchi, & Klimesh, 2000). These methods have mainly focused on prioritization and transmission of 2D images only. On the other hand, most of the solutions developed in academia and industry for 3D reconstruction utilize either computationally intensive algorithms or require human intervention (see the review by Bernardini & Rushmeier, 2002), both of which are not desirable for planetary science applications. It can take several minutes to execute these algorithms using the onboard processors of a rover assuming that the storage capacity is sufficiently large. In addition to the semiautomated reconstruction techniques which require human intervention, there are also some automated methods that combine image-based and geometry-based rendering techniques for photorealistic renderings of scenes. For example, Debevec, Taylor, and Malik (1996) present an approach for modeling and rendering architectural scenes from a set of 2D images. However, a preliminary scene model must be available and the geometry should not be complex to implement this approach. Pollefeys, Koch, Vergauwen, and Van Gool (2000) present another approach for automated 3D reconstruction of architectural scenes from a sequence of 2D images captured by a single camera. This approach can deal with unknown camera settings and no prior knowledge about the scene is necessary to build 3D models, but the mesh quality of the model may not be sufficient to render finer geometrical details such as the textures of a rock.

We have developed efficient methods for 3D registration and integration of range images captured by a planetary rover. In addition, we propose a simple protocol for progressive transmission of the integrated data to Earth. In our approach, we perform the registration of range images

at the remote site. Following the registration, we organize the registered data points spatially using an octree for progressive transmission. Then, we reconstruct multiresolution 3D models of the imaged object from the progressively transmitted octree data on Earth. Finally, we display these models to scientists and mission engineers through a new visualization table and a haptic device in a multimodal virtual environment for scientific analysis. The major benefits of our approach include a) a reduction in the amount of data to be processed by the rover, b) an efficient protocol for progressive transmission of 3D data from a remote planet to Earth, and c) new visualization tools for interactive exploration of the transmitted data in multimodal virtual environments.

2 Data Acquisition

A series of stereo images is produced by navigating a mobile platform, equipped with a stereo head, around a rock in JPL-MarsYard (http://marsyard.jpl.nasa.gov/). The stereo images of the rock surface from different viewpoints are captured by adjusting the camera parameters such as the pitch and yaw, focal length, gain, aperture, and frame rate. Following the acquisition of rock images, range coordinates are extracted from the stereo images using a triangulation method. In order to reduce the volume of range data to be processed, the stereo images are sub-sampled before the range extraction step. Each point in the range map contains a 3D coordinate and an RGB color.

3 Registration

The range maps are aligned (i.e., registered) into a common coordinate system to form a single 3D volumetric representation (see Figure 2). Mathematically speaking, registration of two range maps refers to finding the best 3D transformation that minimizes the distance between them. The Interactive Closest Point (ICP) algorithm is used to register two overlapping range maps (Besl & McKay, 1992; Chen & Medioni, 1992). The alignment of two range maps is achieved in three consecutive steps using the ICP algorithm.

```
repeat
    identify the corresponding points of
    range maps
    compute the optimal transformation
    error = compute the total distance
    between the corresponding range points
    num_iterations++;
    if (num_iterations > iteration_threshold)
    break
until error < error_threshold
```

First, the corresponding points in two range maps are identified. For this purpose, the closest range points in two range maps are determined based on Euclidean distance (i.e., point to point correspondences). One can also determine the correspondences by computing the closest surface points, also known as point–plane correspondences. While the point–plane approach is more robust, it is computationally more expensive than the point–point approach (Pulli, 1999). It also requires a surface representation of one of the range scans to find the closest pairs. After finding the corresponding points using the point–point approach, the rigid transformation that reduces the distance error between two range maps is calculated. Hence, for the given two sets of range points, the problem is to find the optimal transformation (i.e., a rotation matrix and a translation vector) that minimizes the distance error between them in a least squares sense. The closed form solution of this transformation is obtained using orthonormal matrices (Horn, Hilden, & Negahdaripour, 1988). The solution is iterated until the distance error is less than a predefined threshold value. For computational efficiency, we set an upper bound for the number of iterations. In other words, we force the registration process to stop after a certain number of iterations is reached even if the error is not less than the predefined error threshold. This upper bound is set to a small number in order to reduce the number of computations and to prevent the system from entering into an infinite loop. Obviously, this approach may cause high registration errors if the upper bound for the number of iterations is reached before the error criterion is satisfied. However, if the overlapping range scans are initially aligned such that they are sufficiently close to each other before the regis-

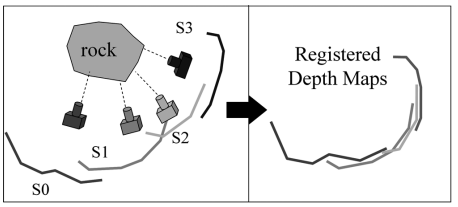


Figure 3. *Registration of depth maps.*

tration process starts, the number of iterations and the total registration error can be reduced significantly. In our application, the initial transformations of the range scans can be estimated using the poses of the rover. The pose of a rover (i.e., position and orientation) can be measured using its onboard sensors such as wheel encoders and gyros. (Note that the pose estimations using sensor readings are subject to errors as the rover moves. Hence, they cannot be used alone to register range maps.)

Although ICP algorithm solves the pairwise registration problem, it is known that global registration is not optimal when pairwise registration is used sequentially to align multiple range maps. While many techniques have been proposed to solve the global registration problem in the past (see the review by Bernardini & Rushmeier, 2002; and Pulli, 1999), this is still an active area of research. We developed a suboptimal solution by grouping the range maps in a hierarchical manner. For example, if we have four range maps as shown in Figure 3 (S0, S1, S2, S3), their registration is performed in two consecutive steps (assume that S0 is the reference range map and the initial transformations are already applied to S1, S2, and S3 to bring the overlapping scans close to each other):

```
step 1: register S1 to S0 and S3 to S2
         using ICP algorithm
        save the transformation of S1
        save the transformation of S3
        group S0 and S1 to "S0S1"
        group S2 and S3 to "S2S3"
step 2: register "S2S3" to "S0S1" using
         ICP algorithm
        ungroup "S2S3" to S2 and S3
        save the transformation of S2
        update the transformation of S3
```

In our registration experiments, we have observed that this way of registering range maps yields better results than the sequential registration approach (S0 to S1, S2 to S1, and S3 to S2 in the above example), but the registration errors still accumulate at each hierarchical level.

4 Integration

Following the registration, we integrate the aligned depth maps into a single unified 3D representation. Integration of range maps is a computationally intensive and difficult problem because range data may contain noise and holes and the sampling density may not be sufficient for extracting a surface model from it. Although novel approaches have been proposed in the past (Turk & Levoy, 1994; Curless & Levoy, 1996), they do not completely automate the integration process. Even if we assume that this process can be automated, it is almost impossible to handle the integration problem using the onboard computational resources of today's rovers due to the large memory requirements.

In our approach, the registered scans are merged to a single volumetric representation at the remote site without constructing a surface model. Then, the integrated data is transmitted to Earth for isosurface extraction. The transmission of the integrated volumetric data to Earth as one big package is not an ideal solution. Even if the data is compressed, the file size will be too large for the direct transmission. Since the transmission bandwidth, rate, and interval between Earth and Mars are limited, data must be transmitted progressively. We use an octree-based subdivision scheme to integrate the range points into a hierarchical volumetric representation for progressive transmission. The main advantage of octree representation over any 3D representation is that it is based on topology and no attempt is made to characterize the underlying geometry of the object (Chen & Huang, 1988).

In an octree representation, the space enclosing a 3D object is subdivided into octants recursively until a predetermined resolution is achieved (see Figure 4). In our implementation, we merge all the range points into a single volumetric representation. We then sub-

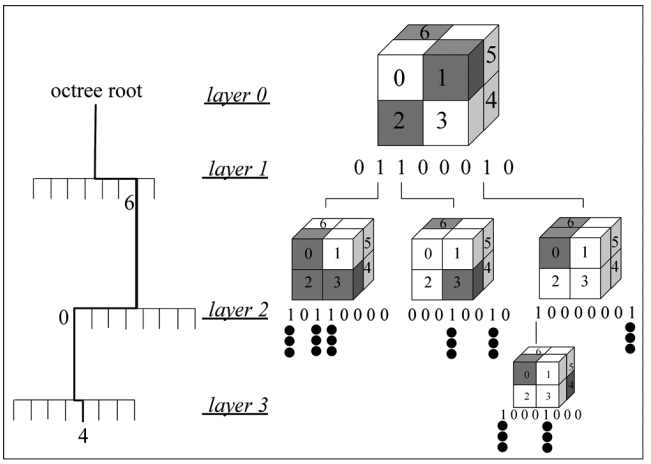


Figure 4. *Our octree-based encoding approach for progressive data transmission. Each node of the tree is represented by a series of 8 bits (one for each children of the node) where "1" indicates that the child contains at least one range point.*

divide the space enclosing the integrated data into eight cubes first and check each partition. If a partition contains more than one range point, we subdivide that partition into eight cubes again and repeat the same process until the maximum number of subdivision levels is reached.

5 Progressive Transmission

The octree representation not only organizes the registered range points coherently to reduce the data size but also enables us to transmit the integrated data progressively. This gives the scientists a rough idea of what they are looking at initially and then allows them to gradually refine it as more data arrives. For transmitting range points progressively using an octree, one can first calculate the average of range coordinates in the occupied octants and then transmit them layer by layer to Earth. Increasingly higher resolution models of the object can be displayed to the scientists as more layers of the octree arrive. In order to reduce the transmission load, we suggest transmitting the center coordinates of the occupied octants instead of the average of the range coordinates in the same octants. The main advantage of this approach is that if a path to each node of the octree is transmitted, then the center coordinates of all nodes along the path can be recovered easily on Earth. Let's say that the path to a node is given by three numbers (6-0-4) for the octree shown in Figure 4. This information tells us that this node is in the third layer of the octree. In addition, it is the fifth octant of the first octant of the seventh octant of the root node. If we know the coordinates of the bounding cube that encloses the object (i.e., the root node), it is now easy to compute the center coordinates of each node along the given path. Moreover, if each node of an octree is encoded as a series of 8 bits, where "1" and "0" represent full and empty children respectively, a path to each node can be extracted automatically from the progressively transmitted bits. For example, if we transmit "0 1 1 0 0 0 1 0" for a node, it becomes obvious that only the second, third, and seventh children of this node contain data and the other partitions are empty (see Figure 4). Naturally, we expect to receive only three new sets of 8 bits in the next transmission cycle. Using this simple encoding scheme, all the branches of the tree can be progressively reconstructed on Earth from the transmitted series of bits. The same octree structure can be used for the transmission of colors as well. The colors are typically represented by intensity values and each intensity value (R, B, G) is in the range of 0 to 255. Hence, each color value requires 3 bytes of storage per range point, which is excessive. To reduce the transmission load, we first average the color values of range points that are clustered in the occupied octants of the octree as implemented by Yemez and Schmitt (2003) and then quantize them to 16 bits (5-6-5) as implemented by Rusinkiewicz and Levoy (2000). As a result of these optimizations, a total of 24 bits (8 bits for each coordinate and 16 bits for each color value) is necessary for storing and transmitting a range point with color information in our application.

6 Transmission Scenario

In our transmission scenario, data arrives on Earth progressively in increasing resolution. Due to the tight

networking constraints of our application (i.e., low bandwidth and transmission rate), we transmit the encoded paths and quantized colors only. We then generate 3D multiresolution models of the imaged object from the progressively transmitted data for visualization in virtual environments. According to our reconstruction scenario, the center coordinates of the occupied octants are first recovered from the transmitted octree paths and then normal vectors are estimated from these coordinates on Earth for isosurface extraction. Since the original range coordinates and surface normals are not transmitted and the surface normals are estimated from the quantized range coordinates (i.e., the center coordinates of the occupied octants), the proposed transmission protocol leads to the reconstruction of less accurate surface models of the object at the early stages of the transmission. However, the errors due to the quantization of range coordinates and the estimation of surface normals from the quantized coordinates become less significant as more layers of the octree are transmitted to Earth in a progressive manner (see Figure 5).

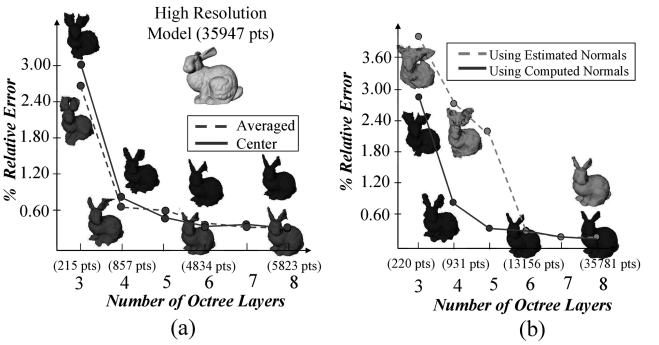


Figure 5. *a) The range scans of the Stanford Bunny are used to compare the 3D surface models reconstructed from (1) the averaged coordinates of range points that fall into the same octants of an octree and (2) the center coordinates of those octants. A total of ten subsampled range scans are used to reconstruct 3D triangular surface models of the bunny in various resolutions from the integrated range data stored in different layers of the octree. The bunnies (colored red in the original) below the graph lines in the figure are reconstructed by using the average coordinate of range points clustered in the occupied octants. The bunnies (colored blue in the original) above the graph lines are reconstructed by using the center coordinates of the same octants. The mean error in each reconstruction, relative to the high fidelity model of the bunny (35,947 points) and based on the Hausdorff distances, is calculated using public domain mesh comparison software (Aspert, Santa-Cruz, & Ebrahimi, 2002). b) A comparison of the 3D bunny models that are reconstructed (1) by using the normal vectors computed directly (dark bunnies, colored blue in the original) versus (2) by using the normal vectors estimated from the range coordinates (light bunnies, colored green in the original). In both approaches, the surface models are reconstructed from the decoded range coordinates.*

7 Reconstruction

Multiresolution and 3D surface models of a rock located at JPL MarsYard are reconstructed from the range images captured by a mobile platform (Figure 6). For enabling touch interactions with the virtual models of rock, a point-based haptic rendering algorithm (Ho, Basdogan, & Srinivasan, 1999) is implemented (see the review of haptic rendering concepts in Srinivasan & Basdogan, 1997; and Basdogan & Srinivasan, 2002). In order to construct multiresolution surface models of the rock in the form of triangular meshes, the Marching Cubes (MC) algorithm is used (Lorenson & Cline, 1987). The basic principle behind the MC algorithm is to subdivide space into a series of small cubes (also known as voxels) and then march through each of the cubes testing the corner points and replacing the cube with an appropriate set of polygons. The polygons generated through this process construct a surface model that approximates the volumetric data set. The MC algorithm combines simplicity with high speed and works almost entirely on lookup tables. To implement the MC algorithm, we first compute the signed distance of each voxel point to the transmitted coordinates. We use a positive (negative) sign for the voxel points that are outside (inside) the rock surface. In a typical implementation of the MC algorithm, a voxel point is internal (external) to a surface if the dot product of the vector from the voxel point to the closest point on the surface (i.e., distance vector) is smaller (greater) than zero. Since we transmit the quantized range coordinates and colors only, we have no explicit surface representation of the rock to decide if the voxel point is internal or external to the rock surface. To make this decision, a group of N

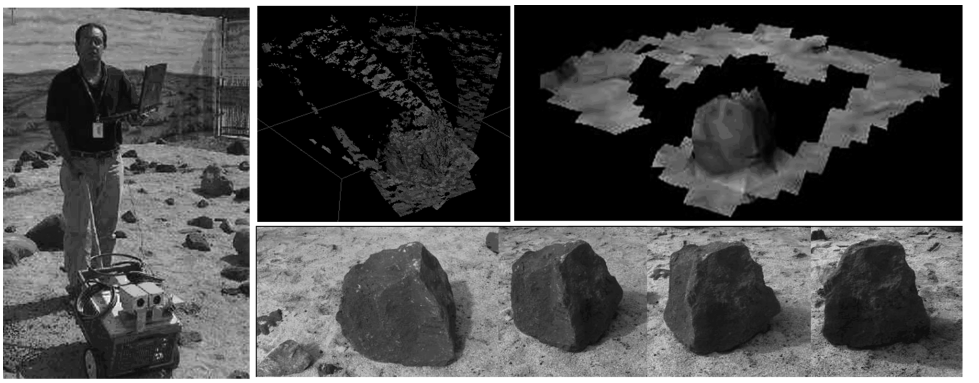


Figure 6. *A mobile cart equipped with a computer and stereo cameras is used to capture multiple views of a rock for 3D reconstruction. Multiresolution and 3D tangible models of the rock are reconstructed from the four range images shown on the left using an octree structure. The original images are subsampled to reduce the number of range points. The 3D model (547 vertices and 3356 triangles) displayed on the right is the highest resolution representation of the rock extracted from the eighth layer of the octree.*

neighboring range points in close proximity to the voxel point and weighted according to their distance to it is used. If the dot product of the distance vector from the voxel point to the neighboring range point and the normal vector at the neighboring range point is greater (smaller) than zero, the voxel point is assumed to be outside (inside) the rock surface and the signed distance is set to a positive (negative) value. We repeat this "in/out" check between the voxel point and all its neighboring range points to decide if the voxel point is internal or external to the rock surface. If the number of "in" ("out") counts is more than the number of "out" ("in") counts, then the voxel point is assumed to be inside (outside) the surface of the rock. The pseudocode of this procedure is given below:

```
float compute_signed_distance (voxel vox)
{
  for j = 1 : N // N is an odd number
  {
   //vox.ngh[ ]contains a list of neighboring range
   points
   sign = vox.ngh[j]. distVec. dot (vox.ngh[j].
   normal)
   if (sign < = 0)
  {
num_inside++;
inside_dist = inside_dist + vox.ngh[j].distVec->length();
  }
   else
  {
  num_outside++;
 outside_dist = outside_dist + vox.ngh[j].distVec-
 >length();
  }
  if (num_inside > num_outside)
  signedDist = – (inside_dist/num_inside);
  else
  signedDist = (outside_dist/num_outside);
  }
 return signedDist;
}
```

As it is obvious from the pseudocode, these computations cannot be done without having the normal vectors of range points. We use the technique proposed by Hoppe, DeRose, Duchamp, McDonald, and Stuetzle (1992) to estimate the normal vectors from the recovered range coordinates. First, we fit a tangent plane to each range point using its close neighbors. The principal component analysis and a minimal spanning tree (MST) are used to determine the direction of the normal vector for the tangent plane. We then calculate the signed distance from each voxel point to the nearest tangent plane. Once the normal vectors and the signed distances are known, it is relatively easy to generate an isosurface passing through the locations where the distance function takes zero values. To eliminate the jagged edges that the MC algorithm typically generates, a simple blending algorithm is used (see Figure 6). We repeat this reconstruction process for each transmitted layer of the octree to generate multiresolution surface models of the rock.

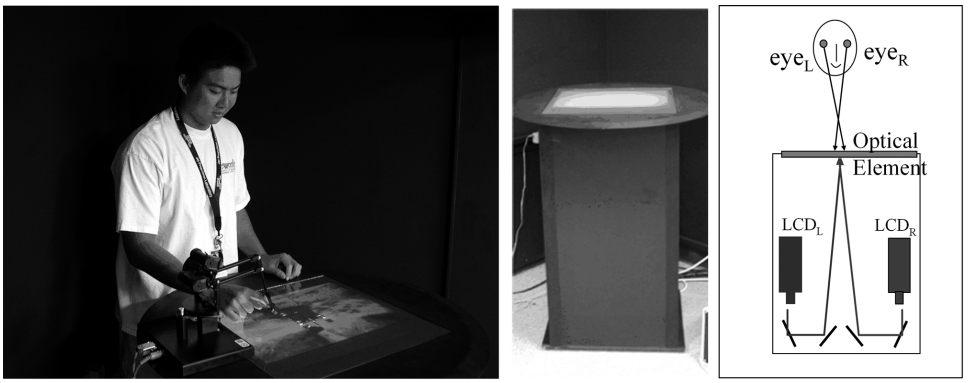


Figure 7. *Our visualization setup includes a new projection table for stereo visualization of 3D objects without using any special eyewear and a haptic device for force feedback.*

8 Visualization of Reconstructed Rocks

We developed a visualization setup for displaying the reconstructed 3D surface models of Martian soil and rocks in a multimodal virtual environment through the use of a new autostereoscopic display and a haptic device (Basdogan et al., 2002). Most of the earlier multimodal visualization systems that integrate stereo displays and haptic devices have utilized polarized or shutter glasses for stereo vision (Chen, Young, Anderson, Jiang, & Nagata, 1998; Veldkamp, Turner, Gunn, & Stevenson, 1998; Brederson, Ikits, Johnson, & Hansen, 2000). Our setup allows a user to touch, feel, and manipulate a virtual object through a haptic device while he/she visualizes its surface model in stereo without using any special eyewear. The setup (see Figure 7) includes a new projection table for autostereoscopic visualization of 3D objects and a PHANToM haptic device (available from Sensable Technologies Inc.) for simulating touch interactions with them. The table was designed to enable a mission engineer or a scientist to stand in front of it and visualize the planetary data more naturally than the existing visualization systems.

Our autostereoscopic display system consists of two LCD projectors (one for each eye) and mirrors housed in a rectangular enclosure, topped with a glass plate (see Figure 7). A graphics card with a dual ported output is used to feed the projectors. Each LCD projector reflects its image off a pair of mirrors, casting the image onto the holographic optical element (HOE) at the tabletop, creating left and right viewing zones. HOEs model the properties of conventional optical elements (e.g., lenses) by holographic methods. Thus, a HOE contains no image information, but serves to diffract rays of light modulated elsewhere (Trayner & Orr, 1996). Our HOE contains two recorded regions (1.25×24 in.) located symmetrically at the center of the plate. These regions were recorded on a thin film using techniques similar to those used for making holograms, but unlike a typical hologram, our HOE diffracts the light beam coming from the projectors to reconstruct image stripes for left and right viewing zones. In our system, there are six image stripes per viewing zone, alternating so that each kind of stripe passes only the light from its corresponding projector while blocking the light from the other projector. This allows each eye to see a different view of the object, creating a stereo image without the use of polarized glasses.

9 Autostereoscopic Display of Martian Rocks

Autostereoscopic visualization is a relatively new technology that allows the viewer to see 3D images without the aid of special glasses (Perlin, Paxia, & Kollin, 2000). Various types of autostereoscopic display systems have been developed during the last decade (see

the review by Hale, 1997). Among these, parallax displays are the most common autostereoscopic displays. Parallax displays consist of a surface covered with display elements that can emit light of varying intensity in different directions. Our autostereoscopic display table uses a holographic optical element to create a parallax effect.

The reconstructed 3D surface model of the rock is rendered in two steps for autostereoscopic visualization: first, two separate views of the model for left and right eyes are created. Then, the way each view must be displayed to a user is determined in order to create the desired stereoscopic effect (see TGS Open Inventor Documentation; Grinberg & Siegel, 1994; McAllister, 2002). In order to create two slightly different views of the same image in our system, two perspective cameras are utilized (one for each eye). To create a stereo effect, we take advantage of the asymmetric view frustum concept. An asymmetric frustum is obtained by skewing the view frustum of the perspective cameras such that the view volume still contains the object and the viewing direction is kept unchanged. To achieve this effect, we first create two perspective cameras in our scene graph (they are placed at eye level, separated by the distance of left and right eyes, and point towards the plate as shown in Figure 8a) and then apply a shear transform to the objects displayed through these cameras. Here, we should point out that orienting the cameras such that they point to the center of the object does not create a proper stereo effect (Hodges & McAllister, 1990). This approach has major problems: the point of focus has to be updated as the position of the displayed object is changed in the scene or the orientation of one view becomes different from the orientation of the other view.

The necessary perspectives for stereo imaging are formulated in terms of a monoscopic camera and a shear transformation that skewed the camera's view frustum. The general form of the shear matrix used in our rendering is:

$$\mathbf{S} = \begin{bmatrix} I + \tan(\phi)(\vec{v}^T\vec{w}) & 0 \\ -(\tilde{Q}\cdot\vec{v})\vec{w} & 1 \end{bmatrix} \quad (1)$$

where $\tilde{Q}$ is a point on the shear plane, $\vec{v}$ is a unit vector perpendicular to the shear plane, $\vec{w}$ is a unit vector in the direction of the shear and perpendicular to $\vec{v}$ (i.e., $\vec{w}\cdot\vec{v} = 0$) and ϕ is the angle of the shear (see Figure 8a). A shear transformation matrix that shifts the point of perspective from $[0\ 0\ r_z]$ to $[r_x\ r_y\ r_z]$ while not shifting the x-y plane at $z = 0$ is used. This matrix is obtained by using the following values and relations (see Arvo, 1991): $\tilde{Q} = [0\ 0\ 0]$, $\vec{v} = [0\ 0\ 1]$, $\vec{w} = (r_x/\sqrt{r_x^2+r_y^2}, -r_y/\sqrt{r_x^2+r_y^2}, 0)$ $\tan(\phi) = \sqrt{r_x^2+r_y^2}/r_z$, where, $[r_x\ r_y\ r_z]$ is the location of the eyes relative to the table top (see Figure 8a; note that r_y has a negative value if the coordinate frame is defined as shown in the figure and the value of r_x is different for each eye resulting in a different shear vector and transform for each camera). After inserting the necessary values, the following matrix is obtained:

$$\mathbf{S}_{camera} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \frac{r_x}{r_z} & -\frac{r_y}{r_z} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (2)$$

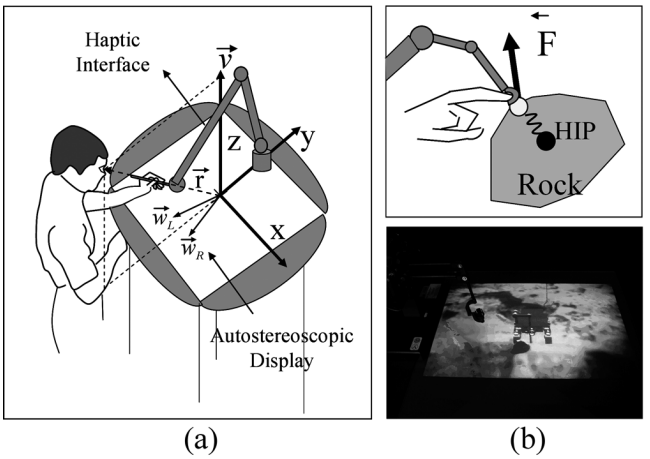


Figure 8. *a) The coordinate frame and the relative distances that are defined in the computation of the shear matrix. b) In our simulations, the tip point of the end-effector (Haptic Interface point, HIP) is used to interact with the 3D model of a rock surface and the surface around the rock. The kinematics of the 3-link haptic arm is mapped to the kinematics of a 4-link PUMA arm to simulate the effect of touching and exploring Martian rocks using the manipulator arm of a planetary rover.*

For implementation, we start with a projection matrix for a standard perspective camera located at $l = r_z$:

$$\mathbf{P}_{monoscopic}(l = r_z) = \begin{bmatrix} r_z & 0 & 0 & 0 \\ 0 & r_z & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & r_z \end{bmatrix} \quad (3)$$

We then apply the shear transformation to obtain the sheared projection matrix as:

$$\mathbf{P}_{shear} = \begin{bmatrix} r_z & 0 & 0 & 0 \\ 0 & r_z & 0 & 0 \\ r_x & -r_y & 0 & -1 \\ 0 & 0 & 0 & r_z \end{bmatrix} \quad (4)$$

In order to implement this concept using Open Inventor Graphics Toolkit, two perspective cameras (one for each eye) located at $[0\ 0\ r_z]$ and pointing to the origin are generated. The shear transform matrices associated with the left and right eye views are applied to identical copies of the object. This process generates two slightly different and skewed views of the same object. The sheared copies of the object are then displayed in two separate windows that are constructed by dividing the screen into two equal halves.

10 Touching and Feeling Martian Rocks and Soil

Previous scientific investigations on Martian soil have been conducted using a manipulator arm in the Viking missions and by spinning the wheels of the Sojourner rover in the Pathfinder mission. In the more recent Mars missions, the rovers collected rock samples and performed more sophisticated scientific analysis to investigate the geological features of the Martian rocks and soil. In our discussions with the field geologists at JPL, it became obvious that displaying the shape and surface properties of Martian rocks through a haptic device would be a valuable addition to our visualization system. We were told that a field geologist can understand a great deal about the composition and formation of rocks through haptic explorations. Using the proposed system, scientists can possibly conduct their field studies on Martian rocks in virtual environments on Earth. In order to display the shape and surface properties of the multiresolution models of the rock through a haptic device, a point-based haptic rendering technique is used (refer to Ho et al., 1999, for the details). As the user manipulates the end-effector of the haptic device, the tip coordinates of the end-effector (also known as Haptic Interface Point, HIP) are digitized via its encoders and the collisions between the HIP and the virtual model of the rock are detected. Once the HIP penetrates into the object, the contact point (also known as Ideal Haptic Interface Point, IHIP) is determined and the penetration vector is calculated as the difference between the IHIP and HIP coordinates. Then, the penetration vector is multiplied by a constant coefficient and displayed to the user through the haptic device (see Figure 8b).

11 Mapping between Visual and Haptic Workspaces

In our visualization setup, the visual and haptic workspaces, defined by the physical dimensions of the visualization table and the haptic device, respectively, are unequal and limited in size. Moving the perspective cameras up or down along the viewing direction to adjust the visual workspace is not feasible because these changes directly influence the shear transformation matrices and result in distorted views of the visualized object (refer to Equation 4). Moreover, the viewing volume of the cameras cannot be altered significantly since their aspect ratio (the ratio of height angle to width angle of the viewing frustum) has to match the aspect ratio (i.e., the ratio of length to width) of the holographic plate so that the displayed objects are in the viewing frustum. In addition to the visualization constraints, the haptic device used in our simulations has a limited physical workspace ($13 \times 18 \times 25$ cm) and sensing resolution which hinder the exploration of very large and very small objects.

We developed a coordinate transformation method that makes the size mismatch between the visual and haptic workspaces transparent to the user of the system (Basdogan et al., 2002). We transform the absolute 3D coordinates of the reconstructed model of the rock to the visual and haptic workspaces independently such that the visual

representation is small enough to be enclosed by the viewing volume of our visualization table while the haptic representation is large enough to be explored with the haptic device. Similar types of transformation techniques have been developed in the past for virtual reality systems that include visual displays and tracking devices (Robinett & Holloway, 1995; Kessler, Bowman, & Hodges, 2000). For implementation, we first multiply the absolute 3D coordinates of the rock surface by the transformation matrices T_1 and T_3 separately such that the object fits into the haptic and visual workspaces respectively (see Figure 9). We then synchronize the cursor movements of the haptic device in the visual and haptic workspaces via the transformation matrix T_2 as shown in Figure 9. Finally, the interaction forces calculated at the absolute coordinate frame are multiplied by a scale factor and the rotation component (R_3) of the transformation matrix T_3 before being displayed to the user. As a result, the forces felt by the user are compatible with the visual position of the cursor on the rock surface during haptic explorations.

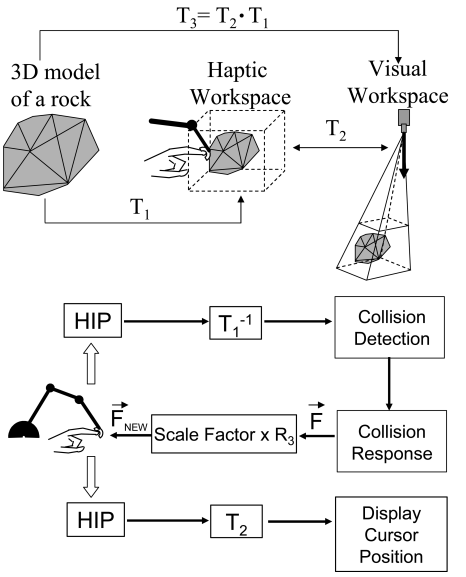


Figure 9. *Transformations that take place between the object space and the haptic and visual workspaces.*

12 Conclusion and Discussion

We presented efficient techniques for the reconstruction of 3D models of Martian soil and rocks from the 2D range images captured by a planetary rover. This is a challenging task to accomplish for a planetary rover using its onboard computational resources only. Since the processing power and storage capacity of a planetary rover are very low, efficient methods are necessary to integrate the acquired stereo images into a unified 3D representation. Following the integration, the data has to be transmitted to Earth for scientific analysis and mission planning. However, the transmission rate and bandwidth between a rover and Earth are very low and progressive transmission techniques are necessary for the safe arrival of data on Earth in an efficient manner.

A hierarchical and group-based registration method was developed for the registration of multiple range scans. This method is computationally less expensive than the existing global registration methods and returns fewer registration errors than the standard sequential registration method. To reduce the registration errors, we suggest that the rover poses are estimated via its onboard sensors at the locations where the stereo images are captured and then used for the initial alignment of range scans before the registration process starts. Following the registration of range scans, the range data was integrated into a single volumetric representation using an octree. In our implementation, we subdivided the occupied octants recursively if they contained more than one data point. One can develop a better subdivision strategy to reduce the number of nodes and branches of the octree by taking into account the distances of range points to the octant centers or the range normals. We have tested a few different subdivision schemes with the Stanford Bunny, but did not observe significant improvements in the mesh quality of reconstructed models. This area may require further investigation.

According to our transmission scenario, the rover transmits the data to Earth for 3D isosurface reconstruction. One can extract an isosurface from the integrated data at the rover site and transmit a compressed surface mesh to Earth instead of transmitting quantized range points. However, the surface extraction process is com-

putationally more expensive than the registration and integration processes. Moreover, attempting to construct a surface model at the rover site from the integrated range data may cause problems since the range data contains noise and gaps. Most of the existing 3D reconstruction techniques are semiautomated and require human intervention to fill these gaps. For these reasons, it is not currently feasible to execute a surface extraction algorithm at the rover site to obtain a 3D surface model of the rover's environment. It is much more efficient to transmit the quantized range coordinates and colors progressively to Earth for 3D reconstruction.

We use an octree structure for efficient quantization and transmission of range coordinates and colors. We consider the center coordinates of the occupied octants as our representative range points for the transmission. In our architecture, transmission of the encoded paths of an octree is sufficient for recovering the quantized range coordinates on Earth. As an alternative to transmitting the quantized center coordinates, Yemez and Schmitt (2003) transmit the average of the range points in the occupied octants. They propose an octree-based transmission protocol to transmit a 3D surface mesh over the internet and render it progressively on the fly. The data transmitted using our protocol is less accurate than their protocol, but our approach is easier to implement and computationally more efficient. For example, Yemez and Schmitt (2003) delete and merge octree nodes for differential encoding, but both operations are computationally more expensive than a simple node insertion into an octree. Moreover, they transmit normal vectors along with the average coordinates for accurate reconstruction of 3D objects at the transmitted site. To reduce the transmission load in our application, we prefer not to transmit the normal vectors. The normal vectors can be successfully estimated from the transmitted coordinates on Earth using the method suggested by Hoppe et al. (1992). We show that the 3D models reconstructed using the estimated surface normals and the ones reconstructed using the original surface normals are not significantly different from each other (see Figure 5b). After the quantization of range coordinates and colors using the proposed octree encoding, 24 bits (8 bits for the coordinates and 16 bits for the color values) are necessary to transmit a data point. For example, only 120 kbits are necessary to transmit a 3D model with 5000 points. With the transfer rates of the Sojourner (40 bits/s) and the new generation rovers (120 kbits/s), this would take 3000 s (not feasible at all) and 1 s (a more reasonable number), respectively.

We also developed a new multimodal visualization system that enables a user to see the virtual models of Martian rocks in stereo without using any special eyewear while touching and exploring their surfaces through a haptic device. Our visualization system uses a holographic optical element with two recorded vertical slits, two projectors, and stereo rendering algorithms to generate an autostereoscopic view of 3D objects. This HOE-based visualization system is relatively cheap and easy to produce, but it can display autostereoscopic models of 3D objects to a single observer only and cannot accommodate arbitrary observer position and orientation. However, space mission planning and scientific visualization of planetary data typically involve multiple people. If multiple viewing zones can be generated on the same holographic plate, mission operators and scientists standing around the table can interact with virtual objects and each other through their personal haptic devices to analyze the transmitted planetary data.

In order to fuse visual and haptic images in our current visualization system, we developed an efficient coordinate transformation technique. This technique enables us to map the absolute 3D coordinates of a virtual object into visual and haptic workspaces using a simple set of matrix transformations. A visual-haptic mismatch frequently occurs in virtual environments when the mapping between visual and haptic workspaces is not one to one. The proposed mapping technique can be easily adapted for other visualization systems that integrate visual and haptic displays for exploration of very large-scale or very small-scale objects. In addition, we have observed that alternative implementations of this mapping could affect our perception of object shapes, which will be further explored in the future. For example, although the haptic representation of the ground surface around the rock was flat (see Figure 6), the surface was perceived to be inclined by the users of the system during demonstrations when the cursor move-

ments were deliberately mapped to follow the visually inclined representation of the ground surface. This observation suggests that multimodal visualization systems such as the one proposed in this study can be, for example, used to augment the perceptual and cognitive skills of a mission operator if the sensory modalities are displayed in certain order and frequency.

Acknowledgments

The research described in this paper was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. It was supported by TMOD, IPN, and CSMISS programs at JPL. We acknowledge the JPL-Machine Vision group for capturing stereo images and converting them to range maps. We are thankful to Prof. Marc Levoy of Stanford University for providing us the scan registration software during our development, which is now available through their website (http://graphics.stanford.edu/software/scanalyze/). We used some parts of the source code to develop our automated registration method and the code that runs on the mobile platform. Bunny scans used in this study are also courtesy of Stanford University. The progress reports submitted by Prof. Isaac Cohen (USC) to JPL on 3D reconstruction and the class notes of Prof. Brian Curless (University of Washington) on 3D photography were helpful to our development. Physical Optics Cooperation (POC), in collaboration with JPL, built the hardware components of the visualization table through a NASA-SBIR grant. Stephen Kupiec (POC) contributed to the development of rendering code for the autostereoscopic visualization. We acknowledge the discussions made with Andrew Kostrzewski of POC on holographic displays. Edward Chow and Larry Bergman of JPL provided management support for the development of the visualization system. Mitchell Lum and Jose Salcedo of JPL helped with the system integration and were involved in the technical discussions on visualization.

References

Arvo, J. (Ed.). (1991). *Graphics Gems II.* (pp. 338–341). Boston: Academic Press.

Aspert, N., Santa-Cruz, D., & Ebrahimi, T. (2002). MESH: Measuring errors between surfaces using the Hausdorff distance. *IEEE International Conference on Multimedia and Expo, 1,* 705–708. Available at http://mesh.berlios.de. Last accepted 8 Sep., 2006.

Basdogan, C., & Srinivasan, M. A. (2002). Haptic rendering in virtual environments. K. Stanney (Ed.), *Handbook of virtual environments* (pp. 117–134). Mahwah, NJ: Erlbaum.

Basdogan, C., Lum, M., Salcedo, J., Chow, E., Kupiec, S., & Kostrzewski, A. (2002). Autostereoscopic and haptic visualization for space exploration and mission design. *Tenth IEEE International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems,* 271–276.

Bernardini, F., & Rushmeier, H. (2002). The 3D model acquisition pipeline. *Computer Graphics Forum*, *21*(2), 149–172.

Besl, P. J., & McKay, N. D. (1992). A method for registration of 3D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *14*(2), 239–256.

Brederson, J. D., Ikits, M., Johnson, C. R., & Hansen, C. D. (2000). The visual haptic workbench. *The Fifth PHANToM User's Group Workshop,* 46–49.

Chen, H., & Huang, T. (1988). A survey of construction and manipulation of octree. *Computer Vision, Graphics, and Image Processing*, *43*, 409–431.

Chen, T., Young, P., Anderson, D., Jiang, Y., & Nagata, S. (1998). Development of a Stereoscopic Haptic Acoustic Real-Time Computer (SHARC). *SPIE, 3295,* 171–178.

Chen, Y., & Medioni, G. (1992). Object modeling by registration of multiple depth maps. *Image and Vision Computing*, *10*(3), 145–155.

Curless, B., & Levoy, M. (1996). A volumetric method for building complex models from depth maps. *SIGGRAPH,* 303–312.

Debevec, P., Taylor, C. J., & Malik, J. (1996). Modeling and rendering architecture from photographs: A hybrid geometry- and image-based approach. *SIGGRAPH,* 11–20.

Drew, K. G., Bowman, D. A., & Hodges, L. F. (2000). The simple virtual environment library—An extensible framework for building VE applications. *Presence, 9*(2), 187–208.

Grinberg, V. S., & Siegel, M. (1994). Geometry of binocular imaging. *SPIE Conference on Electronic Imaging, Stereoscopic Displays, and Virtual Reality Systems*, *2177*, 56–65.

Hale, M. (1997). Autostereoscopic displays and computer graphics. *SIGGRAPH*, *31*(2), 58–62.

Ho, C., Basdogan, C., & Srinivasan, M. A. (1999). An efficient haptic rendering technique for displaying 3D polyhe-

dral objects and their surface details in virtual environments. *Presence*, *8*(5), 477–491.

Hodges, L. F., & McAllisten, D. F. (1990). Rotation algorithm artifacts in stereoscopic images. *Optical Engineering*, *29*(8), 973–976.

Hoppe, H., DeRose, T., Duchamp, T., McDonald, J., & Stuetzle, W. (1992). Surface reconstruction from unorganized points. *SIGGRAPH,* 71–78.

Horn, B. K., Hilden, H., & Negahdaripour, S. (1988). Closed-form solution of absolute orientation using orthonormal matrices. *Journal of the Optical Society of America A*, *5*(7), 1127–1135.

Kessler, G. D., Bowman, D. A., & Hodges, L. F. (2000). The simple virtual environment library, an extensible framework for building VE applications. *Presence*, *9*(2), 187–208.

Kiely, A., Manduchi, R., & Klimesh, M. (2000, June). Maximizing science return: Data compression and onboard science processing. *JPL-TMOD Technology and Science Program News,* 12.

Lorenson, W. E., & Cline, H. E. (1987). Marching cubes: A high resolution 3D surface construction algorithm. *SIGGRAPH*, *21*(4), 163–169.

Manduchi, R., Castano, B., & Cohen, I. (2001, January). Onboard science processing and 3-D modeling for smarter communications. *JPL-TMOD Technology and Science Program News,* 13.

McAllister, D. F. (2002). 3D displays. *Wiley Encyclopedia on Imaging,* pp. 1327–1344.

Perlin, K., Paxia, S., & Kollin, J. S. (2000). An autostereoscopic display. *SIGGRAPH*, *33*(3), 319–326.

Pollefeys, M., Koch, R., Vergauwen, M., & Van Gool, L. (2000). Automated reconstruction of 3D scenes from sequences of images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *55*(4), 251–267.

Pulli, K. (1999). Multiview registration for large data sets. *Second International Conference on 3D Digital Imaging and Modeling (3DIM'99),* 160–168.

Robinett, W., & Holloway, R. (1995). The visual display transformation for virtual reality. *Presence*, *4*(1), 123.

Rusinkiewicz, S., & Levoy, M. (2000). QSplat: A multiresolution point rendering system for large meshes. *SIGGRAPH,* 343–352.

Srinivasan, M. A., & Basdogan, C. (1997) Haptics in virtual environments: Taxonomy, research status, and challenges. *Computers and Graphics Journal*, *21*(4), 393–404.

TGS Open Inventor 6.0 (2006). Stereo viewing (Chap. 10). *User's Guide*. Available at www.tgs.com/support/oiv_doc/index.htm.

Trayner, D., & Orr, E. (1996). Autostereoscopic display using holographic optical elements. *SPIE, 2653,* 65–74.

Turk, G., & Levoy, M. (1994). Zippered polygon meshes from depth maps. *SIGGRAPH,* 311–318.

Veldkamp, P., Turner, G., Gunn, C., & Stevenson, D. (1998). Incorporating haptics into mining industry applications. *The Third PHANToM User's Group Workshop.*

Vergauwen, M., Pollefeys, M., & Van Gool, L. (2003). A stereo vision system for support of planetary surface exploration. *Machine Vision and Applications*, *14*(1), 514.

Yemez, Y., & Schmitt, F. (2003). Multilevel representation and transmission of real objects with progressive octree particles. *IEEE Transactions on Visualization and Computer Graphics*, *9*(4), 551–569.