RANGE IMAGES

A *range image* or *point cloud* is any set of numerical data that can be interpreted as a set of densely sampled three-dimensional (3-D) surface points. Range images have the following properties:

- They are digitized using a wide variety of different sensing technologies: from satellite-based imaging radars that generate terrain maps to laser-based body scanners for custom uniform tailoring to interferometric scanners inspecting the flatness of video tape surface.
- They are processed digitally by algorithms closely related to those used with audio and video data owing to their digital signal properties and computer-aided design (CAD) data owing to their geometric properties.
- They are used in many different 3-D applications, from industrial reverse engineering to inspection to medical shape digitization.

The term *range image* is used as the title for this article instead of *point cloud* to contrast the data content of these geometric signals with gray-scale and color video images, which are familiar to most electrical engineers. In fact, many range images are now cast into formats that have few similarities to video images.

The picture on a black-and-white television is a gray-scale image, where each point sampled in the image possesses a shade of gray ranging from black to white. Whereas most people are quite accustomed to thinking of the shades of gray as representing the amount of light at a point, you may equally well imagine the shades of gray representing the distance from a point on the screen to the nearest surface being viewed (as measured along the same optical ray that would correspond to the intensity measurement in the optical case). For each gray-scale image you view, a corresponding range image represents the 3-D point structure in that image. Figure 1 shows an example of a registered pair of intensity and range images of a man walking into a room.

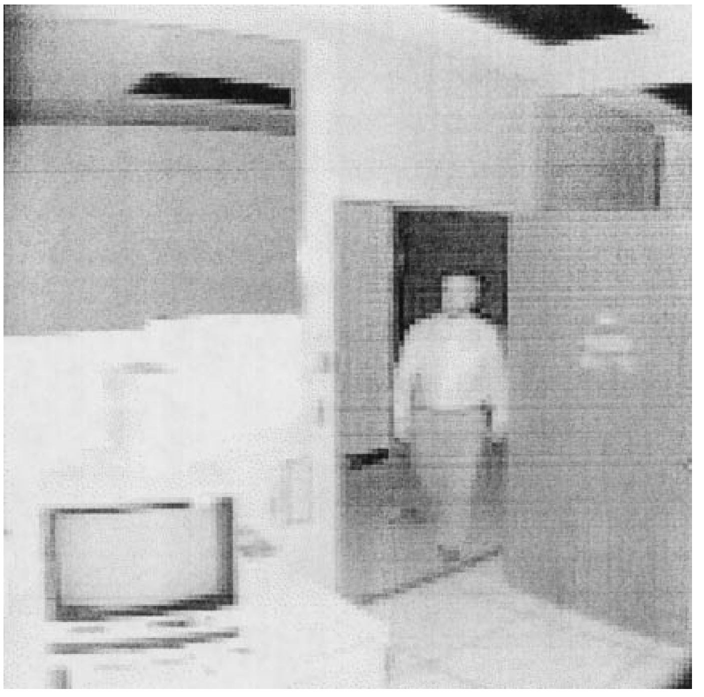
Any color image viewed on a computer screen or a television screen can be represented as a rectangular matrix of three-valued samples (or pixels) where the three values are the intensity of red, green, and blue (r, g, b). Now imagine that in addition to knowing the color at each pixel in an image that you also know the 3-D coordinates (x, y, z) of each surface point visible at each pixel relative to some coordinate system. That collection of 3-D spatial points is a "color range image" or "XYZ/RGB image" because the knowledge of the range, or distance, from the viewer is implied by the knowledge of the 3-D coordinates of the sample points and the 3-D position of the viewer and the image plane.

Another property of images that leads us to the term range image is dense sampling. Whereas people can identify musical passages with as few as three or four musical notes, no one wants to be asked to identify visual scenes or objects based on either the brightness or color visible at three or four pinhole dots poked in an opaque piece of black cardboard or on four (x, y, z) points. Just as images are not perceived as real-world scenes until thousands of closely spaced sample points are available, range images are not perceived, or processed, as real-world surfaces without a similar number of points.

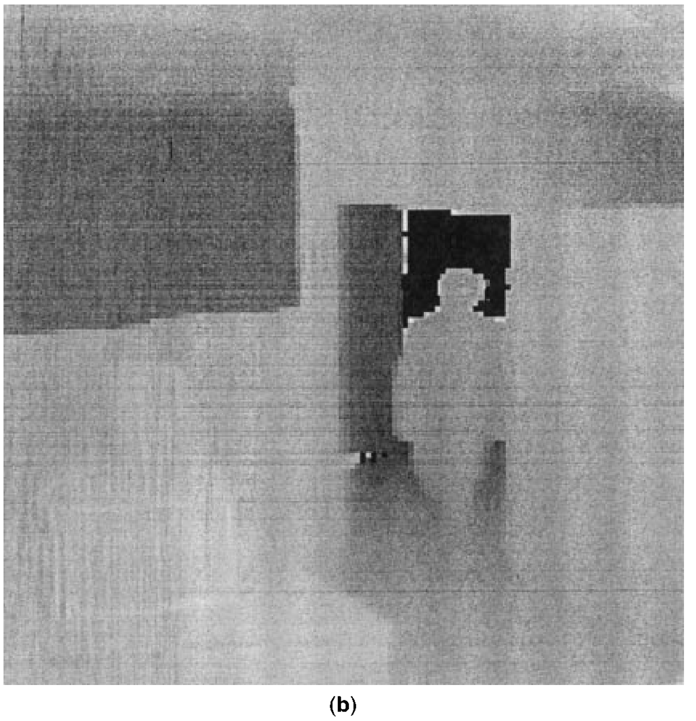
The geometric concept of a range image is known by other names in different disciplines: cloud of points, point cloud, range picture, range map, depth image, depth map, 3-D image, 2.5-D image, 3-D scan, 2.5-D scan, digital terrain map (DTM), topographic map, surface contours, surface profiles, *xyz* point list, surface height map, contour map, and z-buffer, among others. The practice of digitizing and processing range images takes place in quite diverse fields.

HISTORY

Computer vision research began in the mid 1960s. After the first attempts to get a computer to analyze and interpret the output of a video camera, it became apparent that it was difficult to interpret the 3-D world based only on luminance or color. In the 1970s, researchers began looking for methods to acquire 3-D range image information directly to substitute for or to accompany video information. By 1980, commercial technology had evolved to the point that a shape replicator that acquired hundreds of thousands of points on surfaces in a scene in a matter of about 10 s was built. Then that shape was milled directly into wax or plaster to create a solid object copy. By the mid 1980s, imaging laser radars were assisting autonomous robot vehicles and allowing industrial robots to determine grip points on visible parts automatically in a jumbled bin of parts problems. During this time, photogrammetry equipment also became increasingly computerized and automated. Coordinate measuring machine vendors made optical measurement probes available to complement the standard touch probes. By the early 1990s, video-rate 3-D scanners demonstrated the ability to measure 10,000,000 densely spaced points per second even though no 3-D application software even now is close to being capable of processing input at that rate.



(a)



(b)

Figure 1. (a) 128 × 128 gray-scale intensity image of "man walking into room." This image looks very similar to a black and white television picture although it was created by recording the amplitude of the return laser signal from a range-imaging laser-radar system. (b) A completely registered 128 × 128 range image representing 16,384 separate 3-D points. The darker areas represent points that are farther away and the lighter areas represent points closer to the range-imaging sensor. The time required to digitize these two images was 0.6 s.

Although the technology has continued to improve as it has evolved, commercial success still falls short of the immense potential. A combination of factors has tended to be responsible: (1) the cost, reliability, and repeatibility of the range sensing hardware; (2) the cost of the computers to process the data at sufficiently fast rates; and (3) the level of automation in end-user application software, which is often the primary limiting factor.

APPLICATIONS

A list of some applications for 3-D range imaging systems follows:

- Reverse engineering—the creation of 3-D engineering data from functioning physical parts.
- Industrial design using prototypes—the acceleration of the design process by assisting the flow of data from a physical design prototype to conventional CAD data.
- Rapid prototyping from models—the creation of STL (stereolithography) files directly from point data from parts.
- Part replication—any method or process that produces a copy of a 3-D shape from an original.
- Finite element modeling—the creation of shell or solid elements from a physical part.
- Robotic bin picking—sensing 3-D shapes in a bin of parts and determining gripper positions for a robot.
- Robotic assembly—sensing the state of an assembly process to guide a robotic assembly process.
- Robotic navigation—sensing the environment of a mobile robot or moving end effector in order to plan paths.
- Apparel/automated tailoring—digitizing body shapes to extract tailoring information.
- Museum artifact digitization—digitizing the color and shape of rare fragile artifacts once so that they can be examined repeatedly by others.
- Inspection/validation/metrology—measuring points on parts and comparing them to expected points based on "golden" correct models.
- Cartography/surveying—digitizing 3-D terrain for geographic information systems (GIS).
- As-built architecture assessment—digitizing and modeling buildings, plants, ships, and the like.
- Medical sensing—sensing the shape, volume, surface areas of body surfaces for quantitative monitoring of biological geometry.
- Virtual environments—graphics is easy, but modeling is difficult so using models derived from range images can lighten the work load in assembling virtual graphical environments.
- Modeling for animation—film and television artists that use special effects can benefit from ease of digitizing complex shapes of prototype models.
- Modeling for entertainment—game programmers can also benefit from the ability to digitize existing character sculptures.
- Portrait sculpture—several scanner companies have provided systems for this market only to find it smaller than expected.
- Automated object recognition—in limited domains, practical recognition systems are possible.

All these applications share a common need for 3-D coordinates on the surfaces of existing physical objects.

Figure 2 shows a rendering of an example range image data set of an automotive door inner panel consisting of over 3,000,000 3-D data points (a 36 megabyte "range image"). This type of data would be encountered in a reverse engineering application if a physical prototype had to be con-

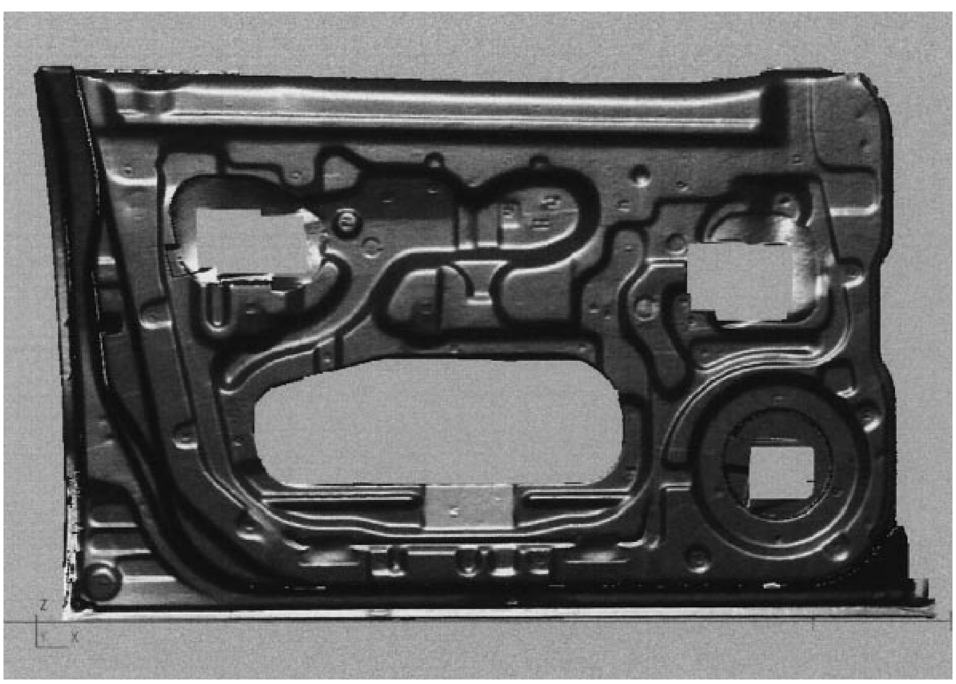


Figure 2. A rendering of a range image data set consisting of 3 million points. Data sets such as this are encountered in reverse engineering as well as inspection/validation applications.

structed first to predict and compensate for sheet metal "spring-back" effects. After the data is digitized, a geometric model is constructed to represent the object as it will be constructed. This type of data can also be encountered in an inspection/validation application where the digitized data is compared to the ideal geometric model to see where subtle shape differences are occurring.

METHODS FOR ACQUIRING RANGE IMAGES

A large variety of devices have been developed to generate range images, and new types of devices seem to be announced on a continuing basis. The simplest device is the tape measure, which can be used to measure the distances from points on surfaces to three orthogonal planes. Because this device is error-prone and labor-intensive, other devices are preferred in real applications. The most common ingredients of commercially available range imaging devices are

- Photon emitters (such as laser diodes),
- Photon detectors (such as charge-coupled device (CCD) cameras),
- Motion devices, controllers and/or sensors,
- Computers with fast processors and fast graphics,
- *A priori* knowledge about scanned surfaces.

An example of *a priori* knowledge would be a precondition that any objects being digitized by a scanner are painted white and do not have a shiny or mirrorlike surfaces.

As an introduction to the scope of range imaging devices, we provide an outline of device categories.

Active Optical Triangulation

Triangulation can be understood using the law of sines from trigonometry. If A, B, C are the three points of a triangle, the angles of the triangle $\theta(A)$, $\theta(B)$, $\theta(C)$ are related to the lengths of the sides $d(A, B)$, $d(B, C)$, $d(C, A)$ by the following equation:

$$\frac{\sin[\theta(A)]}{d(B,C)} = \frac{\sin[\theta(B)]}{d(C,A)} = \frac{\sin[\theta(C)]}{d(A,B)} \tag{1}$$

where $d(\;)$ is the normal euclidean distance between two points. If the length $d(B, C)$ of one side (the baseline) is known, and the angles $\theta(B)$ and $\theta(C)$ are known, then you can compute everything about the triangle including the angle $\theta(A)$ and the lengths of the other two sides $d(A, B)$ and $d(A, C)$.

In practice, a camera is calibrated in 3-D relative to a reference coordinate system. If N 3-D points $\mathbf{x}_i = [x_i\; y_i\; z_i\; 1]$ are imaged at the corresponding 2-D points $\mathbf{u}_i = [u_i\; v_i\; 1]$, the camera calibration matrix is a 3×4 matrix $[C]$ such that

$$[C]\mathbf{x}_i \approx \lambda_i \mathbf{u}_i \tag{2}$$

where λ_i is a scalar value that represents the "distance" along the 3-D ray corresponding to the vector $\mathbf{u}_i$ that you must travel on to meet the point $\mathbf{x}_i$ (can be measured independently in any convenient coordinate system). After the camera is calibrated, the equation that maps a pixel $\mathbf{u} = [u\; v\; 1]$ to a 3-D ray is given by

$$[x\, y\, z] = \mathbf{a} + \lambda[A]\mathbf{u} \tag{3}$$

where $[C] = [A^{-1} \mid -A^{-1}\mathbf{a}]$ and A is 3×3 and $\mathbf{a}$ is 3×1.

For precision measurements, nonlinear lens distortion must be modeled and compensated even though high-quality lenses may produce distortion only on the order of a single pixel on a typical 640×480 image grid. If you were to build a scanner using a CCD camera and a conventional lens using only linear methods, you would find that when you digitize a flat plane guaranteed flat down to optical wavelengths, your range data will still exhibit curvature in both directions. The simplest useful nonlinear model for camera optics maps a measured pixel $[u\; v]$ to a compensated pixel $[u'\; v']$

$$[u'\, v'] = [u\, v] - k((u - u_0)^2 + (v - v_0)^2)[(u - u_0)(v - v_0)] \tag{4}$$

where k is the quadratic radial distortion coefficient (nominally on the order of 5×10^{-8} for a National Television Standards Committee (NTSC) CCD camera) and $[u_0\; v_0]$ is the optical axis image center, which is usually within about 10–20 pixels of the expected center.

Single Point Projection. In this case, a ray of light, typically from a laser diode, is projected on a surface. The ray of light from the light source must be calibrated to determine the equation of the 3-D ray $\mathbf{r}(t) = \mathbf{r}_0 + t\mathbf{n}$. A peak of intensity is detected in the camera's image plane. The neighborhood of the peak is usually processed to estimate a subpixel position of the ideal peak. The peak's position is compensated for lens distortion, and the 3-D ray for that $[u\; v]$ is computed as previously. The measured point is then computed as the midpoint of closest approach between the two 3-D rays.

Single Light-Stripe Projection. In this very popular case, a plane of light, typically from a laser diode equipped with a cylindrical lens to act as a beam spreader, is projected on a surface. A bright contour appears in the image plane of the camera if the surface is within range. For each scan line of the camera, a separate peak is detected. The subpixel position estimate can be generated only within the given scan line as each scan line acts as a separate 3-D point detector. The peak from each scan line is compensated for lens distortion, mapped to a 3-D ray, which then intersects the precalibrated plane of light. The intersection of the ray and the plane is the 3-D detected point. For a 640×480 CCD camera, at most 480 separate 3-D point estimates can be generated for each digitized frame. Even low-cost systems can generate over 1000 points per second.

Circle Projection. Circle of light projection has some advantages in that properties of the neighborhood of a point can be detected.

Binary Pattern Projection. If half of a scene is illuminated and half not, the transition from bright to dark is like the contour in light-stripe projection except now we know that the 3-D points corresponding to the bright pixels are on one side of the 3-D plane represented by the light/dark transition, and the 3-D points corresponding to the dark points are on the other. By subdividing each set of imaged points with half light and half dark projections recursively, you can compute depth resolved to 2^N depth levels using N binary light/dark masks. Liquid crystal display (LCD) projectors are often used for creating the masks. The bit sequence of bright/dark pixel readings places each pixel in 3-D along its $[u\ v]$ ray intersecting at the appropriate plane determined by the bit sequence.

Multiple Light-Stripe Projection. It is possible to compute 3-D coordinates for multiple light stripes in one image if there is some way to keep track of which 3-D plane from which 3-D light stripe is being viewed at each pixel in the image. One way to do this is with a color camera and color-coded stripes.

Grid Projection. Just as you can do multiple stripes, you can use multiple stripes from orthogonal directions projecting a grid of lines on a scene. More sophisticated schemes are required to keep track of which plane of light corresponds to which contour in a digitized image.

Random Texture Projection. Just as you can project multiple stripes on a scene, you can project multiple points in a pattern or in a random texture of bright points projected on a scene.

Structured Pattern Projection. It is possible to come up with a wide variety of light/dark patterns so that when they are projected on a scene, the local point neighborhoods can be disambiguated so as to reliably imply the 3-D ray or 3-D plane relevant to that point.

Passive Optical Triangulation. When two or more feature-rich images of the same scene can be digitized, it is possible to use passive optical triangulation methods to compute depth wherever image features can be matched. Photogrammetry systems and binocular or trinocular stereo systems fall in this category.

Multiple Silhouette Extrusion Intersection. Some objects have shapes that can be very well characterized by capturing two-dimensional (2-D) silhouettes from numerous angles, creating 3-D solids of extrusions and then intersecting all the solids to create the final shape.

Imaging Optical Radars

The basic time/range equation for radars is

$$v\tau = 2r = \text{Round-trip distance} \tag{5}$$

where v is the speed of signal propagation, r is the distance of a reflecting object surface, and τ is the transit time of the signal traveling from the radar transmitter to the reflecting object and back to the radar receiver. For imaging laser radars, the unknown scene parameters at a reflecting point are (1) the range r, (2) the surface reflection coefficient ρ, and (3) the angle $\theta = \cos^{-1}(\hat{n} \cdot \hat{l})$ between the visible surface normal $\hat{n}$ and the direction $\hat{l}$ of the radar beam. Ignoring atmospheric attenuation and scattering losses for close range laser radars, all other relevant physical parameters can be lumped into a single function $K(t)$, which depends only on the radar transmitter/receiver hardware, such as antennas gain factors. The received power $P(t)$ is

$$P(t, \theta, \rho, r) = K(t - \tau)\rho \cos\theta / r^2 \tag{6}$$

Note that for coherent laser delivery of radiant energy, the denominator is the distanced squared, not to the fourth power as in conventional radio frequency radar. There are at least three basic methods to determine range using radar techniques.

Time of Flight. In time of flight (TOF) methods, a pulse of optical energy is emitted at time t_1 and a peak of energy is detected at time t_2 so that $\tau = t_2 - t_1$.

Amplitude Modulation/Phase Difference. Rather than sending out a short pulse, waiting for an echo, and measuring transit time, a laser can be amplitude modulated (AM) by varying the drive current of a laser diode at a frequency $f_{\text{AM}} = c/\lambda_{\text{AM}}$. An electronic phase detector measures the phase difference $\Delta\phi$ in radians between the transmitted signal and the received signal to get the range:

$$r(\Delta\phi) = c\Delta\phi / 4\pi f_{\text{AM}} = \lambda_{\text{AM}}\Delta\phi / 4\pi \tag{7}$$

Because relative phase differences are determined only modulo 2π, the range to a point is determined only within a range ambiguity interval r_{amb}. In the absence of any ambiguity-resolving mechanisms, the usable depth of field of an AM laser radar is the ambiguity interval: $L_r = r_{\text{amb}} = c/2f_{\text{AM}} = \lambda_{\text{AM}}/2$, which is divided into $2^{N_{\text{bits}}}$ range levels where N_{bits} is the number of bits of quantization at the output of the phase detector. Finer depth resolution and smaller ambiguity intervals result from using higher modulating frequencies. Multiple modulation frequencies have been used simultaneously to achieve better depth resolution and large ambiguity intervals simultaneously.

Frequency Modulation. The optical frequency of a laser diode can also be tuned thermally by modulating the diode drive current. If the transmitted optical frequency is repetitively swept linearly between $\nu \pm \Delta\nu/2$ to create a total frequency deviation of $\Delta\nu$ during the period $1/f_{\text{m}}$ (f_{m} is the linear sweep modulation frequency), the reflected return signal can be mixed coherently with a reference signal at the detector to create a beat frequency f_b signal that depends on the range to the object r. This detection process is known as frequency modulated (FM) coherent heterodyne detection. Range is proportional to the beat frequency in an FM continuous wave (CW) radar: $r(f_{\text{b}}) = cf_{\text{b}}/4f_{\text{m}}\Delta\nu$. One method for measuring the beat frequency is counting the number of zero-crossings N_{b} of the beat signal during a ramp of the linear sweep frequency modulation. This zero-crossing count must satisfy the relationship $2N_{\text{b}} = \lfloor f_{\text{b}}/f_{\text{m}} \rfloor$, which yields the range equation $r(N_{\text{b}}) = cN_{\text{b}}/2\Delta\nu$. The range values in this method are determined to within $\delta r = \pm c/4\Delta\nu$ since N_{b} must be an integer. The maximum range should satisfy the constraint that $r_{\max} \ll c/f_{\text{m}}$. Because it has been difficult to ensure the exact optical frequency deviation $\Delta\nu$ of a laser diode, it is possible to measure range indirectly by comparing the N_{b} value with a known reference count N_{ref} for an accurately known reference distance r_{ref} using the relationship $r(N_{\text{b}}) = N_{\text{b}}r_{\text{ref}}/N_{\text{ref}}$.

Focus Methods

The Gauss thin lens law states that a thin lens of focal length f focuses a point light source in a scene at a distance z from the center of the lens onto a focal plane behind the lens at a distance w from the lens center:

$$\frac{1}{w} + \frac{1}{z} = \frac{1}{f} \tag{8}$$

The thin lens law may be expressed as $z(w) = wf/(w - f)$ to show the optimal range for focus at distance w or $w(z) = zf/(z - f)$ to show what distance of the focal plane is needed to focus points of depth z. It is easy to see that as w varies from its minimum value of $w_{\min} = f$ to some maximum value $w_{\max}$, points at depths from $z = \infty$ to the corresponding minimum value $z_{\min} = w_{\max}f/(w_{\max} - f)$ are focused.

A camera lens has a finite aperture of diameter D, and light passing through a finite aperture always experiences diffraction and blurring. The radius of the blur of a point at range z is a minimum when the point is in focus at $w = w(z)$. The blur increases as w varies away from $w(z)$, or similarly, as the point distance z varies away from $z(w)$ (in either direction). If a point blur is modeled as a 2-D Gaussian intensity distribution of diameter σ for a fixed focal plane distance w, the range to that point is given as

$$z_{\pm}(\sigma) = \frac{wf}{w - f \pm \sigma F} \tag{9}$$

where $F = f/D$, the so-called f-number of the lens aperture. If the blur characterized by σ is small enough to maintain adequate focus, the depth of field L_z of the camera in which all points are in adequate focus is given by

$$L_z = z_{-}(\sigma) - z_{+}(\sigma) = \frac{2wf\sigma F}{(w - f)^2 - \sigma^2 F^2} \tag{10}$$

The depth of field increases with increasing f-number F.

An assortment of ingenious methods using lens focus have been developed for producing range images. Whereas focus methods are not usually as accurate as triangulation methods, focus methods can sometimes be implemented at lower cost or higher data rates. Focus methods are also well suited for microscope measurements of 3-D structures in transparent biological samples.

Moire Methods/Spatial Phase Modulation

A moire pattern is a low-spatial-frequency interference pattern created when two gratings with regularly spaced patterns of higher spatial frequency are superimposed on one another. Mathematically, the interference pattern $A(x)$ from two patterns A_1, A_2 is

$$A(x) = A_1\{1 + m_1 \cos[\omega_1 x + \phi_1(x)]\} \cdot A_2[1 + m_2 \cos(\omega_2 x + \phi_2(x)] \tag{11}$$

where the A_i are amplitudes, the m_i are modulation indices (usually about $\frac{1}{2}$), the ω_i are spatial frequencies, and the $\phi_i(x)$ are spatial phases. When this signal is low-pass filtered (LPF) (blurred), only the difference frequency and constant terms are passed:

$$\begin{aligned} A'(x) &= \text{LPF}[A(x)] \\ &= A_1 A_2\{1 + m_1 m_2 \cos[(\omega_1 - \omega_2)x + \phi_1(x) - \phi_2(x)]\} \end{aligned} \tag{12}$$

For equal spatial frequencies, only the phase difference term remains. In moire sensors, surface depth information is encoded in and recovered from the phase difference term. There are numerous moire measurement techniques, but one method has received a significant amount of attention is multiple phase shift moire.

In multiple phase shift moire, N separate images are digitized. The nth frame ($n = 1, 2, \ldots, N$) of image data $I_n^{\text{obj}}(x, y)$ for the object surface is acquired with a grating of pitch P_{p} at the phase shift location $(n - 1)P_{\text{p}}/N$, which corresponds to the phase $\phi_n = (n - 1)(2\pi/N)$. The effective projected pitch at the object is $P_0 = 1/p_0$. The camera and the projector are separated by an angle θ. It is also assumed that N frames of reference images $I_n^{\text{ref}}(x, y)$ for a flat reference surface have been acquired during a preliminary calibration procedure. The object surface images and the reference images are modeled as being of the form

$$I_n^{\text{obj}} = a(x, y) + b(x, y) \cos[\phi_{\text{obj}}(x, y) + \phi_n] \tag{13}$$

$$I_n^{\text{ref}} = a(x, y) + b(x, y) \cos[\phi_{\text{ref}}(x, y) + \phi_n] \tag{14}$$

The reflected ambient light level in the images is $a(x, y)$, the reflected fringe contrast is $b(x, y)$, $\phi_{\text{obj}}(x, y)$ is the phase function of the first image of the object surface prior to grating shifts, and $\phi_{\text{ref}}(x, y)$ is the phase function of the first image of the reference surface prior to grating shifts. After the N image frames have been acquired, the averaged object phase image $\Phi_{\text{obj}}(x, y)$ is computed:

$$\tan \Phi_{\text{obj}}(x, y) = \frac{\sum_{n=1}^{N} I_n^{\text{obj}} \sin(2\pi n/N)}{\sum_{n=1}^{N} I_n^{\text{obj}} \cos(2\pi n/N)} \tag{15}$$

The object phase image is then unwrapped to yield a smooth phase image Φ_{obj}^s without any rapid 2π transitions: $\Phi_{\text{obj}}^s = \Phi_{\text{obj}} + 2\pi m$. It is also assumed that the N image frames of the reference surface were similarly processed to yield a phase image for the reference surface $\Phi_{\text{ref}}(x, y)$ and the unwrapped smooth phase image $\Phi_{\text{ref}}^s = \Phi_{\text{ref}} + 2\pi m$. The phase difference image $\Delta\Phi^s(x, y)$ is computed as the image difference: $\Delta\Phi^s(x, y) = \Phi_{\text{obj}}^s(x, y) - \Phi_{\text{ref}}^s(x, y)$. If the rays of light are approximately parallel when they impinge on the object and reference surfaces, the range image of the smooth surface is given approximately by

$$z(x, y) = \frac{p_0}{\tan\theta} \frac{\Delta\Phi^s(x, y)}{2\pi} \tag{16}$$

Touch Methods and Motion Devices

Imagine a rectangular table with rails running the length of the table on both sides. On these rails, mount a movable fixed-height "bridge" with a rail of its own running the breadth of the bridge. On the bridge rail, mount a counterbalanced vertical linear-translation stage that slides back and forth on the rail as well as moves up and down. If a precision ball-screw is attached to each axis with a motor (steppers or servos) and an optical encoder, you have a cartesian XYZ robot. Servos are commonly used in the higher-quality equipment, whereas steppers are found on inexpensive equipment. On the vertical stage, attach an extension arm and at the end of the arm place a small switch that closes when the slightest force is encountered. The device is now a coordinate measuring machine (CMM) with a touch probe. The point measurement process consists of moving the touch probe either manually or under motor control until it touches a point on a surface. This closes the switch, activating an electronic readout of the *X,Y,Z* encoder values. The encoder values are converted to millimeters, a prerecorded zero offset is subtracted from the reading, and an (x, y, z) coordinate is written to a display and to a file. Repeating the process thousands of times creates a range image.

Coordinate Measuring Machines. CMMs have been used since the 1960s and are the primary tool in the metrology/inspection/validation field. This description was oversimplified in that an automated motorized 2 degree-of-freedom joint is usually attached to the end of the vertical arm, and the touch switch is located on the end of an extension rod attached to this 2 degree-of-freedom joint. Thus, calibration is more involved than a simple prerecorded offset owing to the additional degrees of freedom, but the principles are the same. The data rate for CMMs is usually under 1 point per second, but recent systems can achieve 10 to 20 points/s. The table is usually a large slab of granite. The accuracy of upscale systems is measured in microns.

Multiple Revolute-Joint Arms. Instead of a cartesian robotic device, imagine an N degrees-of-freedom revolute-joint robot arm with optical encoders or potentiometers on each joint where $N = 5$ is typical. Reading out the N values, subtracting calibrated offsets, multiplying by calibrated factors, and combining the transformation matrices for each joint yields the position of a touch probe attached to the end of the Nth arm. Several portable CMMs are available based on this approach. With the correct counterbalancing, an operator can "scribble" on a surface with the probe tip and capture a point every T milliseconds to create a "scribble scan" range image.

Mechanical CMMs with Optical Probes. A very popular and powerful combination is mounting an optical range finding device, such as a point or light-stripe triangulation subsystem, on the end of a CMM arm instead of a touch probe that requires physical contact. The mechanical subsystem of such a combination provides large working volumes and ease of use, whereas the noncontact optical measurement subsystem provides high speed.

Optical-Arm Light-Emitting Diode Touch Probes. Imagine a rigid touch probe wand with LEDs mounted on it such that the LEDs are visible to two or more calibrated cameras on rigid tripods. Each LED creates a separate brightness peak in each camera, which then correspond to known intersecting 3-D rays that determine the 3-D point of each LED. Three LEDs allow you to compute the coordinate frame of the touch probe wand, and four LEDs allow an overdetermined system that provides some averaging. Further system calibration allows the 3-D coordinates of the tip of the touch probe to be computed as long as all the LEDs are visible to two cameras.

This type of system uses passive triangulation to provide the same functionality as the mechanical components of the conventional *xyz* axis CMM or the multiple revolute-joint arm.

Optical-Arm with an Optical Probe. The next logical step is to combine an optical light-stripe probe with the optical linkage of an LED-equipped wand observed by two or more cameras. At the time of this writing, there are no commercial devices of this type.

Stylus Profilometers. A conventional tool used in tribology (the study of friction) is the stylus profilometer. It is very similar in principle to the CMM, but it is a much smaller device.

Scanning Tunneling Microscopes. The atomic scale of matter can be probed using closed loop feedback to control a microscopic probe. The distance between the probe and a surface is maintained based on the amount of tunneling current.

Proximity Sensors

A variety of proximity sensors can be used to make 3-D measurements when mounted on an appropriate motion device.

Magnetic Effects. A low-frequency modulated magnetic field can be used to make 3-D measurements on nonmetallic objects.

Destructive Scanning

Most of the range imaging methods described here use nondestructive noncontact methods of making 3-D point measurements. If the part you are measuring is expendable, scanners can mill away your part one z-layer at a time. Such scanners are based on the same principles of several rapid prototyping machines, but are employed in an opposite manner. Imagine that you have a solid white part with internal structure. First, you embed the object in a block of black wax. You begin removing a layer of the wax and then taking a digital picture and performing edge detection. The camera optics and position are arranged so that the current layer is always in optimal focus and the z-coordinate is read out from an encoder on the z-axis transport system. Although this sort of technique is applicable in a relatively narrow range of applications, the benefits of the internal structure digitization are achieved without using tomography or other more expensive techniques.

Holographic Interferometry

Holographic interferometers use coherent light from laser sources to produce interference patterns due to the optical-frequency phase differences in different optical paths. If two laser beams of the same polarization meet at a surface point $\mathbf{x}$, then the electric fields *add* to create the net electric field:

$$E(\mathbf{x}, t) = E_1 \cos[\omega_1 t - \mathbf{k}_1 \cdot \mathbf{x} + \phi_1(\mathbf{x})] + E_2 \cos[\omega_2 t - \mathbf{k}_2 \cdot \mathbf{x} + \phi_2(\mathbf{x})] \tag{17}$$

where the $\mathbf{k}_i$ are 3-D wave vectors pointing in the propagation directions with magnitude $\|\mathbf{k}_i\| = 2\pi/\lambda_i$, the $\omega_i = \|\mathbf{k}_i\|c$ are the radial optical frequencies, and $\phi_i(\mathbf{x})$ are the optical phases. Because photodetectors respond to the square of the electric field, the detectable irradiance (intensity) is $I(\mathbf{x}, t) = E^2(\mathbf{x}, t)$. Photodetectors themselves act as low-pass filters of the irradiance function I to yield the detectable interference signal

$I'(\mathbf{x}, t) = LPF[I(\mathbf{x}, t)]$, or

$$I'(\mathbf{x},t) = \frac{E_1^2 + E_2^2}{2}\left\{1 + \frac{2E_1E_2}{E_1^2+E_2^2}\cos[\Delta\omega t + \Delta\mathbf{k}\cdot\mathbf{x} + \Delta\phi(\mathbf{x})]\right\} \quad (18)$$

where $\Delta\omega = \omega_1 - \omega_2$ is the difference frequency, $\Delta\mathbf{k} = \mathbf{k}_2 - \mathbf{k}_1$ is the difference wave vector, and $\Delta\phi(\mathbf{x}) = \phi_1 - \phi_2$ is the phase difference. This equation is very similar to the preceding moire equation for $A'(x)$ except that a time varying term is included. Because phase changes are proportional to optical path differences, fractions of an optical wavelength distances can be measured. For equal optical frequencies and equal (wave vector) spatial frequencies, only the phase difference term remains. Just as the z-depth spacing of moire fringes is proportional to the period of grating lines, the z-depth spacing of holographic interference fringes is proportional to the wavelength of the light used. Measured object surfaces must be very flat and smooth.

Other Specialized Methods

Fresnel Diffraction/Talbot Effect. When a coherent light source illuminates a diffraction grating, you can move a white card along the line of light and you can see a high-contrast image of the grating at one depth and no contrast at another depth, and this effect repeats. Between the no contrast and the highest contrast, the contrast of the lines from the grating is directly proportional to the range from the nearest high-contrast position. There are similar effects possible with focussing.

Stereo Electron Microscopy. Electron microscopes generally create gray-scale images of microscopic shapes. Using two registered images and other known calibration parameters allows the use of passive binocular stereo methods in a nonoptical domain.

Computed Tomographic Imaging

Higher-energy levels of radiation can penetrate opaque objects. For example, when x rays (keV photons) or gamma rays (Mev photons) pass through an object, they are attenuated by the material of the object in a predictable manner. Imagine that a point source is positioned at some z-level to radiate an object and a one-dimensional (1-D) array of detectors is used to measure the profile of the amount of radiation that has passed through the object and impinged along a line or arc behind the object. This profile of attenuation values in itself may not seem very interesting. However, if these profiles are digitized for every few degrees of rotation of the object, the combination of the profiles can be transformed into the original density function of the object at that z-layer. By indexing the object through the complete range of z-values for that part, a complete sampled density function is acquired. Thresholding each density image yields 3-D polyline contours, whereas running a 3-D isosurface operator, such as an algorithm known as marching cubes, over the data set produces a complete triangle surface tesselation for the part.

Gamma-Ray Computed Tomographic Imaging. Gamma rays are needed to penetrate large iron structures, such as automotive engine blocks, whereas x rays are satisfactory for most smaller objects. Compared to the laser eye safety concerns of most range-imaging sensors, operators of this type of scanner must be certified in the United States by the Nuclear Regulatory Commission (NRC) because the photon source of this type of scanner is radioactive!

Other types of tomographic imaging used in medicine include Single Photon Emission Computed Tomographic Imaging (SPECT) and Positron Emission Tomography (PET).

Magnetic Resonance Imaging

Radio frequency pulses of electromagnetic energy (radio waves) can interact with the magnetic dipoles of atomic nuclei in a variety of substances, including living tissues. Using the phenomenon known as nuclear magnetic resonance (NMR), it is possible to create 3-D images similar in structure to computed tomographic (CT) images, but containing many rich details without the deleterious effects of x rays. The algorithms for extracting 3-D surface points are a bit more involved than those for processing CT images because magnetic resonance imaging (MRI) does not measure density directly, but similar types of 3-D information are available.

Ultrasonic Imaging

Some types of ultrasonic imaging devices can also produce 3-D measurements of internal surface structures. Most ultrasonic data are much noiser than what is possible with other methods.

Computer Vision Methods

Numerous passive techniques from the computer vision field can successfully yield 3-D measurements of surface points. Many of these fit in the "Shape from X" category: shape from shading, shape from texture, shape from motion, shape from texture, and shape from specularity, among others. Shape or structure from motion approaches have produced the best, most usable data.

RANGE IMAGE QUALITY

How should competitive range imaging devices and methods be compared? Any measuring device is characterized by its measurement *resolution,* its *repeatability,* and its *accuracy.* Whenever measuring devices are operated by human operators, another critical element, known as *reproducibility,* must also be evaluated.

Range resolution or *range precision* is the smallest change in range that a sensor can report. Range repeatability refers to statistical variations as a sensor makes repeated measurements of the exact same distance. Uncalibrated sensors can be repeatable without being accurate. Range accuracy refers to statistical variations as a sensor makes repeated measurements of a known true value. Accuracy should indicate the largest deviation of a measurement from the true value under normal operating conditions. Because range sensors can often improve accuracy by averaging multiple measurements, accuracy should be quoted in the context of the measurement time. Also, accuracy is usually quoted for a given measurement interval (the depth of field). Loss of calibration can occur over time with some systems so it is important to also know the temporal characteristics of a scanner from one hour or day to another.

The method of specifying accuracy may vary across different applications, but an accuracy specification should include one or more of the following for each 3-D direction given N observations:

- the mean absolute error ($\pm\delta_x$, $\pm\delta_y$, $\pm\delta_z$) where $\delta_x = (1/N)\ \Sigma\ |x_i - \mu_x|$ and $\mu_x = (1/N)\ \Sigma\ x_i$ [or $\mu_x = \text{median}(x_i)$];
- the rms (root-mean-square) error ($\pm\sigma_x$, $\pm\sigma_y$, $\pm\sigma_z$) where $\sigma_x^2 = (N - 1)^{-1}\ \Sigma\ (x_i - \mu_x)^2$ and $\mu_x = (1/N)\ \Sigma\ x_i$; or
- the maximum error ($\pm\epsilon_x$, $\pm\epsilon_y$, $\pm\epsilon_z$) where $\epsilon_x = \max_i|x_i - \mu_x|$

Regardless of the measurement error probability distribution, $\delta \le \sigma \le \epsilon$ for each direction.

Range accuracy may be specified as $\pm\sigma$, $\pm 3\sigma$, or $\pm\epsilon$ for any measured point within the working volume V of size $L_x \times L_y \times L_z$. In most industrial applications of 3-D scanners, users want to know the maximum possible deviation $\pm\epsilon$ that a measured point might vary from the "real location."

The preceding parameters specify key spatial properties of the sensor. The pixel dwell time or single point measurement time T is the time required for a single measurement. For multiple points acquired sequentially, the total time to take N_p points is N_pT, and the data rate is $1/T$ points per second or points per hertz.

A figure of merit is sometimes handy for mapping several system parameters to a single number for comparing different systems. An application-independent performance figure of merit F can be defined to measure the quality of range images as well as the spatial and temporal performance of range imaging sensors. The figure of merit F is defined as

$$F = \frac{1}{M\sqrt{T}}\left(\frac{L_xL_yL_z}{\sigma_x\sigma_y\sigma_z}\right)^{1/3} \tag{19}$$

where L_x, L_y, L_z are the lengths of the sides of the bounding box containing a range image or the working volume of a scanner, σ_x, σ_y, σ_z are the rms range accuracies in each direction, T is the measurement time, and M is an associated monetary cost factor.

The dimensions of F may be thought of roughly as the amount of good-quality range data per second, and the higher the number the better, other things being equal. A doubling of the depth-of-field to range-accuracy ratio is reflected by a doubling of the figure of merit. However, a quadrupling of image acquisition speed is required to double the figure of merit. This expresses a bias for accurate measurement over fast measurement but also maintains an invariant figure of merit under internal sensor averaging.

POINT ORDERING ISSUES

The usefulness of the 3-D points to some automated algorithms may vary according to the natural ordering of the points from the scanner. In general, algorithms should be prepared for completely randomly ordered points. However, this is almost never the case. When the data points exhibit structure (e.g., scan lines), that structure can sometimes be used to save processing time or generate higher-level geometric primitives without elaborate processing.

Most range imaging devices fall into one of the following categories:

- No order—randomly ordered points.
- Flying-spot scanner order—forward, backward, or both.
- Raster order—points correspond to some unknown subset of a pixel grid.
- Lace cut order—points are digitized in constant planes and direction alternates across an area.
- Scribble line order—points are acquired from a flexible arm probe.
- Tomographic slice contour order—every contour is closed and has constant z
- Rectangular range image grid—an $L \times M$ matrix of points.
- Cubic density image grid—an $L \times M \times N$ matrix of density values.
- Free area scans—data set of multiple range images in a patchwork.
- Free scan lines—scan lines are not necessary similar to previous or subsequent scans in orientation or other aspects.

Each scanner system can take advantage of its own style of data point generation.

APPLICATION ALGORITHMS

There are numerous applications of 3-D range imaging technology, as listed above, but there are only two fundamental application categories of range-imaging sensors:

- Model construction ("reverse engineering")
- Model comparison ("inspection and validation")

In model construction, the raw data from a scanner is converted into a form that is directly useful. In model comparison, the raw data are compared to a given model and difference information is generated. Both of these main processes typically require two key subprocesses:

- Data alignment for multiple scans
- Noise suppression and outlier removal

In data alignment, the raw data are rotated, translated, and otherwise transformed from the sensor coordinate system into the application coordinate system where it is used from construction or comparison. Noise suppression is usually achieved via some type of smoothing, or low-pass filtering, operation implemented as FIR (finite impulse response) or IIR (infinite impulse response) digital filters.

Two examples of simple functions with real-world applications that might not appear to fit into this categorization are (1) surface area estimation and (2) volume estimation. Both functions require a mathematical interpretation of the point data. Typically, explicit model construction would occur first followed by a sum of triangle areas or subtended volumes. The harder part of the problem is the model construction.

Model Construction

Model construction (also known as reverse engineering) applications can be categorized into those that require computer-aided design/computer aided manufacturing (CAD/CAM) compatible nonuniform (knot) rational B-spline (NURBS) surfaces and those that do not. The mathematics for NURBS surface construction differs for the cases of single polynomial span surfaces (Bezier) and multiple polynomial span surfaces. Another distinction that arises is that some simple modeling systems attempt to do everything with untrimmed surfaces only, whereas most mainstream CAD/CAM tools allow the use of trimmed and untrimmed surfaces.

Surfaces require patch boundaries, and laying out the topology of patch boundaries is the single, most difficult part of surface construction. Generally, surface construction is achieved through a completely manual definition of patch boundaries. Some successes have been achieved in completely automatic processes, but these surfaces are not easily adopted into a conventional design process because no designer would make the same type of surfaces as the automated methods. It is likely that successful near-term practical surface construction efforts will fall into the guided, semiautomatic category, where users provide qualitative toplogical information and the computer determines the quantitative details.

There are plenty of applications where an approximate polygonal mesh will suffice instead of a precise continuous smooth surface description. Because every planar polygon can be decomposed into a set of triangles, and triangles present no structuring limitations, triangle mesh models are the most common and the most efficient to work with. In some circumstances, the advantages of a regular quadrilateral mesh are still compelling.

Beyond polygons and NURBS lie other possibilities, but very few other systems have been developed to the point of practical usage. The following outline summarizes this categorization of model construction techniques:

- NURBS surface structure
  - Bezier/single-span polynomial
  - Multiple-span piecewise polynomial
- NURBS surface type
  - Untrimmed
  - Trimmed
- Surface construction methods
  - Manual
  - Guided, semiautomatic
  - Automatic
- Triangles/polygons
  - Unstructured meshes
  - Regular grids

Other non-NURBS

Model Comparison

Model comparison (also known as inspection, verification, validation) applications accept the definition of a model and raw point data registered in the same coordinate system and then report the deviations of the points from the model surfaces via some graphic means such as color maps or deviation vectors. If acceptability criteria can be formally defined, systems need to partition the geometry into the acceptable areas and the questionable areas. Whereas a wide variety of algorithms are used in model construction, the main computations in all model comparison (and data alignment) applications are *closest distance computations.*

Closest distance computations can be categorized as follows:

- Points to points
- Points to polylines/curves
- Points to polygons/surfaces
- Polylines/curves to polylines/curves
- Polylines/curves to polygons/surfaces
- Polygons/surfaces to polygons/surfaces

Except in the point-to-point case, the minimum distance vectors of valid point comparisons are orthogonal to a curve or surface and approximately orthogonal to a polyline or polygon mesh. The orthogonality condition can be used to ignore points that are not in good correspondence to a model.

The computational complexity of comparing millions of points to millions of other entities might seem totally intractable at first glance. Consider computing the closest point P in a set of N points to a given point q. The single-point computation requires N operations where you compute the distance from q to each point p in P and you take the point with the smallest distance. However, if you execute $N \log N$ operations and sort the points into a tree structure, you can find the closest point in a maximum radius in about $\log N$ operations. Therefore, if you have a set of M points, the total number of operations to compute closest distances is approximately $M \log N + N \log N$ instead of MN, which one might think otherwise. The cost of the faster computation is the additional memory to hold the tree structure. Other schemes that lower computational cost but increase memory usage are also possible.

CONTINUING PROBLEMS

If 3-D optical scanners can digitize points so quickly and if algorithms have been developed for most of the needed functions, why are range-imaging sensors not more commonplace? There are no easy answers to this question, but we'll mention a few issues.

Cost is still a big factor inhibiting widespread deployment of range imaging devices. Most of the commercial systems involve one or more expensive components that make them difficult to produce at low cost. Moreover, the high cost tends to keep the volume small, and vice versa. Active optical triangulation systems are the most common because of relatively simple hardware requirements: a computer, a camera, a laser, and a motion device.

Light source and detector are separated in most optical scanners, with the exception of some radars and some focus methods. This separation causes shadowing effects where no data are collected. In particular, most optical scanners have difficulty sensing in concave regions such as holes. Systems compensate for this by taking multiple scans from different directions to see the shadowed regions. One recently published detailed scan data set required about 70 separate scans (1).

Shiny objects make horrible subjects for most optical scanners unless they are coated with a dull finish. A calcium carbonate spray is a popular finish because it sprays on like a white paint yet wipes off like powder with a dry cloth.

Areas of nonuniform surface reflectance, such as different colors, colored textures, and lettering, are not ideal for range scanning. When a triangulation-based laser scanner crosses a reflectance transition, the centroid of the brightness peak shifts causing a smooth 3-D surface to exhibit small steps.

In fact, whenever an optical scanner of any sort moves its beam across a sharp edge, physical or reflectance, edge effects occur in the range data. Methods have been developed to work around this problem using temporal signal profiles.

Scanners with no internal moving parts are generally more robust against abuse than those with moving parts. Many of the commercial scanning systems have internal moving parts that must be calibrated and maintained.

Spot size of a projected beam is sometimes a key issue. Some very accurate scanners have a spot that is larger than the depth accuracy of the system so that having very precise measurements in all three axes is difficult to achieve.

Some optical scanners have been designed with continuous automatic self-calibration. This is a very desirable feature whenever technically possible.

Many 3-D optical scanners are much heavier than users would prefer. A scanner that weighs less is better, other things being equal.

Most scanner vendors do not provide all the software necessary for a complete application solution. The lack of total solution providers has hindered progress in the past, but the situation is getting better.

Although lasers are an asset, laser speckle can actually reduce the accuracy of some methods, such as the basic light striper, compared to what can be done with a white light system. However, laser diodes are so inexpensive and last so long compared to other light sources that it is difficult to justify using any alternative, especially when white light sources need to be replaced so often owing to filament burnout.

Range-imaging devices are detectors. In detection theory, the probability of false alarms and the probability of misses are analyzed. Most range-imaging devices have a probability of measuring a false point that is too high. Also, the probability of not measuring a given point is also higher than what most users would prefer. When optical scanners are competing with touch probes on metrology devices, for example, these comparatively high probabilities are a disadvantage that some applications cannot overlook, despite any other advantages.

Retro-reflection at concave corners sometimes generates false optical peaks that produce bad point measurements. These outliers must be removed via other means, or scanners must suppress detection directly.

Optical scanners must meet all relevant eye safety standards. Although more optical power may be the right solution for more accuracy, it can also be the wrong solution for practical eye safety reasons.

Although there are research and commercial software solutions to the problem of alignment of multiple range images that work quite well, this task still takes time and user expertise and can never be made too easy. Closure processing, or resolving range image overlap from a cycle or loop of range images, must be done very carefully to create the optimal estimate.

Many users would like to see automatic multiple-range image conversion to a mesh of NURBS surfaces that possess appropriate surface continuity between patches. Eck and Hoppe have proved that such an algorithm is possible (2). However, if it is necessary to integrate automatically created patches with conventionally created surface patches, there is still more research to do.

Although it is becoming less of an issue with faster and faster processors, early applications of range images always suffered from long processing times. No one has yet created industrial application software that can absorb and process all the data at the rate they can be generated.

Although progress has been made in all areas of range data processing, such as automated filtering, segmentation, and edge detection, there are no automated system yet that can solve all the problems all the time, or even most of the problems most of the time.

FUTURE DIRECTIONS

It is difficult to make predictions about the time scales of how events will evolve. Certainly, most researchers in this field in 1980 would not have predicted the current state of the art in 1997. Most predicted that the technology would be much further advanced at this point in time. The future is likely to hold more gradual improvement in all the problem areas already mentioned. One critical process that is taking place is the continual education of prospective users on the realistic expectations of today's technology. Too often people imagine that "starship technology" is here today only to be disappointed rather than sorting out the practical advantages and disadvantages and then working to capitalize on existing advantages.

In the past, there has been a split between measurement technology suppliers and software suppliers. Most organizations that need the technology for their applications are interested in finding single-source providers at reasonable costs. Because of this preference, there is a definite trend for range-imaging system suppliers to provide more and more of the necessary application functionality on the sensor side.

One interesting situation is that the whole premise of people doing reverse engineering is directly contrary to what most people in the mechanical CAD/CAM field have preached during the previous 20 years. Yet while some fall more in line with that approach, more and more people find advantages to being able to take real-world shapes into the computer and manipulate them as if they were a CAD model. At some point in the future, geometry will be geometry, and it will make little difference where it came from or how detailed it is. Unfortunately, the CAD/CAM industry is still dealing with this so it is unlikely to disappear soon. In some ways, 3-D scanning technology is just the next step in an orderly progression that included 1-D bar code scanning and 2-D flat-bed image scanners.

BIBLIOGRAPHY

1. B. Curless and M. Levoy, A volumetric method for building complex models from range images, *ACM Siggraph Proceedings 1996,* pp. 303–312.
2. M. Eck and H. Hoppe, Automatic reconstruction of B-spline surfaces of arbitrary topological type, *ACM Siggraph Proceedings 1996,* pp. 325–334.

Reading List

3D Imaging and Modeling Conference Proceedings, Ottawa: IEEE Press, 1997.

ACM SIGGRAPH Conference Proceedings, 1992–present.

T. Bosch and M. Lescure (Eds.), *Laser Distance Measurements,* Bellingham, WA: SPIE Optical Engineering Press, 1995.

J. Hoschek and W. Dankwort (Eds.), *Reverse Engineering,* Stuttgart: Teubner, 1996.

PAUL J. BESL
Alias|Wavefront, Inc.