

sensors®

YOUR TECHNICAL RESOURCE FOR SENSING, COMMUNICATIONS, AND CONTROL

INTELLIGENT SYSTEMS

Machine Vision: Adding the Third Dimension

Electronic perception technology is a new way to make direct, rather than derived, depth measurements of target objects.



James Spare, Canesta, Inc.

All objects occupy three-dimensional (3D) space. Accurate electronic representations of these objects must therefore measure, calculate, or estimate in three dimensions as well. This requirement applies to all systems that attempt to find, locate, and evaluate 3D objects. Conventional machine vision systems produce 2D images that are appropriate for recording and display. These are intended to be viewed and interpreted by the human eye. The television and movie industries, for example, rely on the capabilities of the human brain to extract contextual information that will create realistic 3D representations from 2D images. When the same images are processed by a computer to recognize, interpret, and track objects, the task becomes immensely difficult because the contextual information is no longer available. Massive increases in processing power and graphics software have revolutionized the ability of computers to *model* 3D objects, but creating an accurate representation of an actual object is fundamentally a data extraction problem.

2D Approaches to the 3D Challenge

The key to “seeing” objects is not to analyze them merely in terms of colors and textures, but rather to segment the scene in terms of real-world objects and their spatial relationships. Several methods are currently used to create usable images from 3D scenes, including 2D simplification, stereo camera, and structured light.

2D simplification, the most common, attempts to overcome the 3D challenge by assuming away or attempting to control the effect of the third dimension. These systems use a lens to focus light reflected by the scene onto a tiny, flat semiconductor chip that contains sensing pixels arranged in a rectangular array. 2D simplification assumes that target objects are at a known distance from the camera and are resting on a flat surface parallel to the surface of the lens.

The stereo camera technique combines two or more 2D images acquired from sensors set at different perspectives from the object. These so-called synthetic 3D representations perform a variety of mathematical operations that convert the 2D images into 3D coordinates.

The structured light approach is similar to the stereo camera, in that it derives 3D data from

Creating an accurate representation of an actual object is fundamentally a data extraction problem.

acquired 2D images. These systems project a specific pattern of light onto the target and calculate the 3D data from differences between the reflections and the original patterns. As a result, structured light systems require substantial postprocessing.

Understanding a 3D Environment Requires 3D Information

In each of the previously described approaches, 2D data are acquired by the sensor for the purpose of creating an artificial, so-called synthetic 3D electronic representation. The 2D data are mathematically converted to create a synthetic third dimension. Because they are unable to acquire actual 3D data, the three synthetic 3D techniques have inherent limitations that affect the performance, speed, and reliability of deployed applications.

For example, reliance on 2D images can make object recognition difficult. Suppose a computer program has identified the outline of an object taken by a standard 2D camera. Each of the real-world objects in Figure 1 (page 34) produces the same camera image, shown on the right. To locate, identify, and differentiate among these objects, the system developer must carefully select the location of the sensor, control the lighting to minimize the introduction of shadows, and be sure to present the objects in an orientation that will enable the sensor to tell the objects apart. Depending on the variety of parts presented and the throughput requirements, such an application may be impossible for a system based on a single 2D camera. Furthermore, the application would likely require complex part orientation methods that tend to increase system cost and complexity, while having a negative effect on system reliability and speed.

Is It Larger or Closer?

A fundamental limitation of 2D simplification systems is their inability to differen-



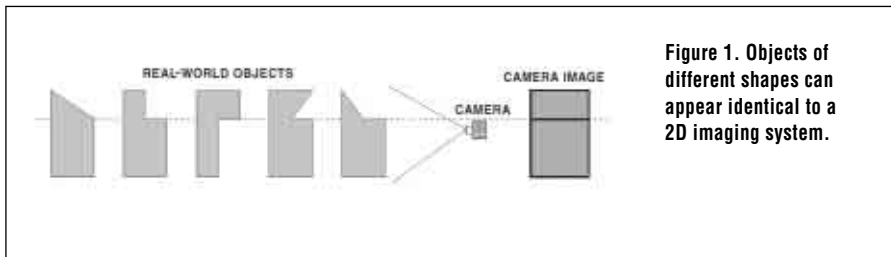
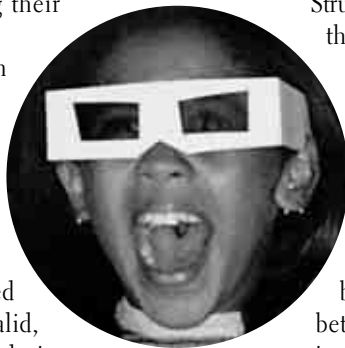


Figure 1. Objects of different shapes can appear identical to a 2D imaging system.

tiate between an object that is larger and another that is simply closer to the camera, making it difficult to determine object size (see Figure 2, page 35). A small object near the camera and a larger object farther from it can create identical images, so it is not possible to know the exact size of objects without knowing their distance from the camera.

To solve this problem with a single conventional 2D sensor, engineers must assume that the object is at a fixed distance from the camera or the system cannot function. To make “fixed distance” assumptions valid, considerable accommodations must be made to precisely control the presentation of the object to the scene. Unfortunately, these accommodations are often associated with increased cost, complexity, and maintenance requirements.



Synchronization and Correspondence of Stereo Systems

In the stereo camera technique, each camera looks at the object from a slightly different vantage point (see Figure 3, page 35). Since both cameras must be looking at the object simultaneously, they must be synchronized. When two frames (one from each camera) are ready, the process of reconciling the pixels in each frame begins. This is typically achieved with a general-purpose processor or inside a specialized chip in an effort to solve the so-called correspondence problem. The algorithm must determine the pair of pixels (one in each frame) that corresponds to the same point in the real world. If it can’t find this correspondence, there will be “holes” in the depth portion of the image. The offset, or disparity, between these two points is then manipulated by triangulation to determine the dis-

tance (or depth) of that point to the camera. This derived method of obtaining 3D depth data requires heavy postprocessing and is prone to error.

Maintenance and Alignment Constraints of Structured Light Systems

Structured light also creates a synthetic 3D representation of an object out of 2D images. These systems send carefully selected patterns of light into the scene and calculate the third dimension by analyzing changes in the patterns. These systems require both perfect synchronization between light and camera, and precise alignment. The maintenance requirements of structured light systems are often quite high.

Throughput Constraints of Scanning Laser Systems

Scanning laser systems use lasers in conjunction with CCD cameras or other sensors, and require the laser to mechanically scan the entire scene line by line and point by point to acquire a single frame. The 3D model must then be constructed one point at a time, which makes the systems too slow for many online applications. In addition, scanning lasers often rely on motors to manage complex rotations and mirrors that add bulk, cost, and maintenance requirements to the overall system.

Electronic Perception

Electronic perception technology is based on a novel method of capturing real-time 3D (depth) images of nearly any object under most ambient lighting conditions: It measures depth by time-of-flight. This technique calculates the difference between the amount of time the light takes to strike different parts

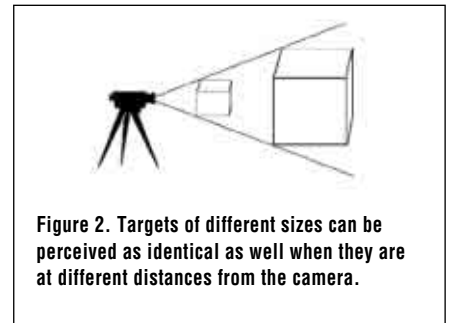


Figure 2. Targets of different sizes can be perceived as identical as well when they are at different distances from the camera.

of an object and then to bounce back to a sensor array. It does not entail multiple cameras and works even in darkness. Ordinary digital video cameras store only intensity values of the target in each pixel of every frame; electronic perception also captures and stores the depth values of each pixel.

An Enabling Technology

The acquisition of depth data from a single sensor greatly simplifies the task of locating the target object in 3D space. The depth data are primary, the result of direct measurements, and therefore do away with having to derive or calculate depth based on a series of other measurements.

Simple. Without the synchronization and correspondence requirements of other methods, electronic perception-based vision systems are inherently simpler to develop, deploy, and maintain. For instance, there are no multiple frames to be synchronized, and the use of primary depth data eliminates the need to take CPU cycles from the application to compute the depth.

Deployable. Electronic perception performs depth measurements inside the chip, without a host processor. Because the electronics can be housed in a small package, it is an ideal candidate for embedded applications where size and weight are important considerations. Single-chip integration greatly reduces the complexity of designing an embedded architecture.

Robust. In many imaging applications,

the system must work under varying lighting conditions. Electronic perception applies various techniques to eliminate the saturation effect of ambient

light, and produces robust range information. For example, the performance of other methods depends on the texture of

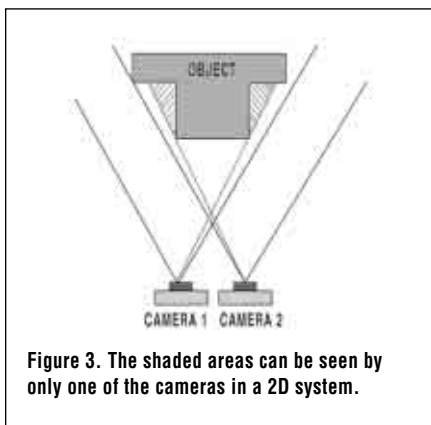


Figure 3. The shaded areas can be seen by only one of the cameras in a 2D system.

the object. If it is too smooth, the correspondence problem is unsolvable and leaves gaps in the depth map. Depth measurement by electronic perception is not affected by target texture or color.

Intelligent Sensors for Object Detection

Electronic perception technology holds promise as the basis of more intelligent sensors for object detection, bin picking, conveyor profiling, materials sorting, and logistics operations. These and other such applications have typically used

2D cameras, structured light, or scanning laser systems, but electronic perception technology may provide advantages in terms of throughput, robustness, line flexibility, maintenance, and speed of deployment. With its foundation in low-cost CMOS fabrication technology, electronic perception provides a new level of price/performance not before seen in the industry, which could yield a new wave of dramatically increased productivity and sophistication. ■

James Spare is Vice President, Product Marketing, Canesta, Inc., San Jose, CA; 408-435-1400, jspare@canesta.com, www.canesta.com.
