The recent convergence of innovations in optics\(^1\) and photonics\(^2\) combined with advances in information technology (IT\(^3\)) has enabled a new world of 3-D sensing technology that is transforming our everyday lives. Whether in the home, in a car, at the mall, in the operating room or on the factory floor, 3-D sensing is ushering in a paradigm shift in how humans interface with their devices and environments.

The success of 3-D sensing technology in consumer electronics, particularly in 3-D gaming systems like the Microsoft Kinect controller for Xbox, popularized its first application—namely, optical gesture recognition.\(^4\) As often seen with disruptive technologies, 3-D sensing has quickly demonstrated commercial viability in numerous industries, such as 3-D imaging sensor chips, device and environmental security, smart lighting and HVAC control, and medical CAD applications—to name but a few. Moreover, increasing interest in the use of 3-D sensing offers tremendous potential as the technology continues to evolve.

This is a large and rapidly growing market. According to recent market research reports, the total value of the gesture recognition and touchless sensing market in 2012 was approximately $2.2 billion, and “is expected to reach $22.04 billion by 2020, [growing] at a double-digit CAGR from 2013 till 2020.\(^5\)” Even the gesture recognition market alone “is estimated to grow from $322.30 million in 2012 to $7.15 billion in 2018.”\(^6\)

**From Stereopsis to Stereoscopy**

Humans (and other animals with two simple eyes) see naturally in three dimensions; immediate, realistic perception of depth and distance has obvious survival value. How do we do it? A light source such as the Sun produces billions of photons of light. These photons reflect off objects in our field of view and enter our eyes. Because of the fixed distance between our two eyes (about 3 inches in humans), two slightly different images are transmitted to the brain, which then uses the apparent change of position (parallax) between the two images to perceive distance and depth in a process called stereopsis.\(^7\)

The earliest 3-D imaging technology fooled the brain into seeing depth in a 2-D image by using similar principles called stereoscopy. In the 1860s, photographers of U.S. Civil War scenes used twin-lens cameras that captured two offset images, much as our eyes do.\(^8\) These twin images were then viewed through an instrument that allowed each eye to see only one image; the spectator’s brain then performs the integration that produces the 3-D effect. A similar approach, only in color, was used in the mid-20\(^{th}\) Century ViewMaster technology.\(^9\)
Early black and white 3-D movies used colored filters to produce a twinned image on a single frame; viewers then wore colored lenses to deliver one image to each eye and voila! — an illusion of three dimensions (and a splitting headache).\textsuperscript{10} More recently, full-color 3-D movies use polarization instead of colors to split the image. Even the path-breaking movie, Avatar, although exploiting advances in optics and computer-generated imagery, ultimately relied on polarized stereoscopy for its 3-D effects.

Beyond Stereoscopy: Three Methods to Capture Dimensional Data

The confluence of advances in photonics with progress in computer power, speed and miniaturization has enabled new approaches to sensing and capturing the third dimension. “Today, there are three common technologies that can acquire 3-D images, each with its own unique strengths and common use cases: stereoscopic vision, structured light pattern, and time of flight (TOF). With the analysis of the 3-D image output from these technologies, gesture-recognition technology becomes a reality.”\textsuperscript{11}

- **Stereoscopy**: Viewing the same scene from two adjacent points of view (eyes or artificial lenses), to visually calculate depth and gauge distance. Requires a brain or computer to integrate the two images.
- **Structured Light Pattern**: Projecting a light pattern (plane, grid, or shapes) at a known angle onto an object/landscape and analyzing the pattern distortions to capture imaging and dimensional information. A computer integrates the data in order to control a system (game, robot, drone) or create an image.
- **Time of Flight**: Methods that measure the time it takes for an object, particle, electromagnetic or acoustic wave to travel a fixed distance through a medium and convert that data into a model or use the data to guide a system such as a game, robot or surveillance camera.

3-D Sensing for Gesture Recognition

Now imagine a computer game where a player swings a golf club in front of a gesture recognition device with 3-D sensors. Illumination sources in the device bathe the player and the surrounding area with invisible, near-infrared light. The light bounces off the player and reflects back to the device. Optical filters screen out stray and ambient light, letting only the near-infrared spectra through to the light sensor, a depth camera\textsuperscript{12}. Interpreting differences in the light patterns bouncing back from different movements of the play, a firmware\textsuperscript{13} chip creates a 3-D map of the player and sends it to the computer game. “The result is a very realistic gaming experience—without wires to trip over or controllers to send flying.”\textsuperscript{14}

3-D Sensing System Components

Regardless of the various methods and technologies used to support 3-D sensing systems, they share the following basic components:\textsuperscript{15}

- **Illumination Sources** — LEDs or laser diodes typically generating infrared or near-infrared light, which is not normally noticeable to users and can be optically modulated to improve the resolution performance of the system. Enabled by Optics 2014 Winner JDSU is a major supplier of illumination sources.
• Controlling Optics — optical lenses and precision beam-shaping components help optimally illuminate the environment and focus reflected light onto the detector surface. A bandpass filter lets only reflected light that matches the illuminating light frequency reach the light sensor, eliminating ambient and stray light that would degrade performance. JDSU is also a major supplier of controlling optics.

• Depth Camera — using a particular sensor or running a stereo algorithm on the frames, a high-performance optical receiver detects the reflected, filtered light, turning it into an electrical signal for processing by the firmware.

• Firmware — very-high-speed custom-designed chips process the received information and convert it into a digital format that can be understood by the end-user application such as video game software.

For 3-D sensing to be leveraged across numerous industries for use in a diverse array of applications, the optical and electronic components used to build them require high performance and reliability. This entails compact, robust and flexible 3-D sensors, along with accurate, smaller, faster and high-efficiency components that allow for high-volume, low-cost manufacturing. Fortunately, the prerequisite standards and levels of sophistication necessary for demanding applications, such as communications networks, consumer electronics, drone surveillance and surgical robotics, etc. have largely been met by companies like JDSU and leading research labs around the globe.16

The other crucial development is the recent coming together of sophisticated optics with improvements in computing technologies and microprocessors that make real-time information processing possible and useful to the end-user. “The problem of manipulating, editing, and storing the measured data was approached at the software level, by developing a very powerful suite of algorithms and programs that import and manipulate the data files, and output them in popular and well-standardized formats.”17

The Role of Gesture Recognition in Natural User Interfaces

“The essence of a natural user interface is invisibility: a user interacts with the technology in a way that hides the number of steps between intent and result. For example, to turn on a light, a complex interface means walking to a wall, finding a switch, and flipping a toggle. A natural interface means merely saying ‘light’ or flicking one’s hand from anywhere in a room.”18

Letting the Machine Use Initiative

Users tend to think of an interface in one direction: they give commands for a machine to perform. The advantage of 3-D sensing as part of a natural user interface is that it is a powerful way to let machines respond to needs before they are expressed without converting the needs into commands. The Human Computer Interface (HCI) works in two directions and the machine can initiate an action independently, without prompting from a human.19

For instance, a physical-fitness video game enabled with 3-D sensing could recognize that a user was using their left arm less than their right arm. The game could then cue the user to address the imbalance, or it could provide a specific set of game conditions that would naturally balance the arm usage. All this would take place without the user’s knowing: the machine would be proactively interacting with the human in an invisible way.
Adding Functionality

3-D sensing will add functionality to existing interfaces rather than eliminate them. An effective user interface provides many different ways to accomplish the same task. For example, a home alarm system typically uses a keyboard as an authentication device to allow entry to any user who knows the code. A 3-D sensing system can recognize a face, a gait, even clothes, and let a user in automatically. But rather than eliminate the keyboard and just use 3-D sensing, it makes more sense to combine the two devices. With 3-D sensing as an add-on to the keyboard, the door will unlock automatically as a user approaches the door; and, if the user has given the door code to a friend, the door will unlock after he or she enters the code on the keyboard.20

Combining Capabilities

“Similarly, 3-D sensing alone can make for a rich experience, but adding additional capabilities with a physical device can heighten the experience exponentially.... Adding 3-D sensing capability and integrating it with other technologies will enhance rather than replace user-interface components.”21

Future Applications

The future of 3-D sensing holds enormous potential for a growing number of sophisticated applications. Thanks to sufficiently mature and ever-evolving 3-D technologies, devices are already becoming affordable for mass-market consumption. Software development kits featuring fast, accurate and low-cost processing functionalities and complex analytics herald a multitude of new products and uses, particularly for mobile devices like smart phones and tablets.22 Beyond that, 3-D sensing technologies are already entering industrial and consumer markets through present and future applications in home automation, biomedical, transportation, surveillance, manufacturing and entertainment (think interactive 3-D movies!).

Today, we are seeing only the beginning of a disruption in interactivity – the future of 3-D vision sensing will span markets and create experiences not yet imagined. Enabled by optics, 3-D sensing is yet another way which photonics brings innovation and improvement into millions of households around the world.

###

1 Optics: the branch of physical science that deals with the properties and behavior of light.

2 Photonics: the applied science of production and manipulation of light that supports such foundations of modern technology as lasers, fiber optics and light amplifiers.

3 Information technology: the applied science behind the computers and telecommunications systems that process, communicate and store data.

4 http://www.finisar.com/blogs/lightspeed/tag/3d-sensing-technology/

5 http://www.marketsandmarkets.com/Market-Reports/touchless-sensing-gesturing-market-369.html

6 http://parker.bus.sfu.ca/bus466bok/index.php/Gesture_Recognition


9 Oral communication with Robert D. Guenther, Adjunct Professor, Dept. of Physics, Duke University, Durham, NC

Firmware: the combination of persistent memory, program code and data stored in a chip

Oral communication with Alexandre Fong, Senior Vice-President, Life Sciences and Instrumentation, Gooch and Housego, Orlando, FL