INDUSTRIAL TECHNOLOGY ADVANCES

Sensification of computing: adding natural sensing and perception capabilities to machines

ACHINTYA K. BHOWMIK

The world of intelligent and interactive systems is undergoing an era of unprecedented innovation and advanced development. With the rapid progress in natural sensing and perceptual computing technologies, devices and machines are increasingly being endowed with the abilities to sense and understand the world, navigate in the environment, and interact with humans in natural ways. Interfaces based on touch sensing and speech recognition are now ubiquitous, and the race is on to the next frontiers of machine intelligence and interactions based on three-dimensional (3D) sensing. In this paper, we review the recent progress in the development of real-time 3D-sensing technologies and their deployment in a new class of interactive and autonomous systems. As an example of a commercially available platform, we describe the Intel® RealSense technologies, as well as its deployments in a new class of interactive and autonomous systems and applications.

I. INTRODUCTION

We have long aspired to build machines that could sense and understand the world, mimicking the human sensation and perceptual processes, and use these capabilities to autonomously navigate in the three-dimensional (3D) world, interact with us in natural and intuitive ways, and assist us in our daily lives at home and at work. Such intelligent and interactive devices and machines have been dreamt up in numerous science fiction novels and movies. This dream is now approaching reality in a number of applications, thanks to the rapid advances in sensing and transduction technologies, powerful and energy-efficient computation and processing hardware, machine vision and speech recognition utilizing artificial intelligence algorithms, besides significant progress in other related fields such as communications, portable energy sources, locomotion and manipulation, etc. The introduction and proliferation of touch-sensing displays in mobile devices have extended human interfaces beyond the traditional keyboard and mouse devices in the recent years, and have enabled direct and natural interactions. Similarly, recent breakthrough developments in speech recognition technologies have enabled voice-based interfaces and interactions. With the advances in new 3D-sensing and processing technologies, various systems and applications are increasingly being

Intel Corporation, Santa Clara, California, USA

Corresponding author:
A.K. Bhowmik
Email: achintya.k.bhowmik@intel.com

introduced in the marketplace with the abilities to "sense", "understand", and "interact" in the 3D world with natural human-interface capabilities [1].

We map and understand the 3D world around us and navigate in it with a robust spatial tracking capability utilizing our visual and vestibular sensing systems. Vision is among the most important of our senses. Equipped with a binocular visual system comprising two eyes and sophisticated visual information-processing pathways in our cortex, we see and understand the world in 3D. Besides the binocular depth perception process, called stereopsis, we also deduce 3D spatial information using other important mechanisms, including a number of monocular depth cues, vestibular motion sensing, auditory perception, etc. The resulting real-time 3D-sensing capabilities are crucial to our seamless understanding, navigation, and interaction in the 3D world, which occur so naturally to us that we take them for granted. However, as we endeavor to implement similar capabilities in machines and devices through hardware, software, and systems technologies, we start to fathom the complexities involved!

In this paper, we review the recent advances in real-time 3D-sensing technologies, and applications in intelligent and interactive devices and systems. As an example of a commercially deployed platform, we present the 3D-sensing and processing hardware and intelligent software capabilities of Intel® RealSense technology, and highlight its recent deployment in a number of applications. Examples include interactive computing systems such as laptops, all-in-one computers, mobile devices, autonomous robots, unmanned

aerial vehicles (UAVs), smart mirrors, as well as a new class of immersive virtual and augmented reality devices. We first briefly review the human 3D-sensing and perceptual system comprising the visual and vestibular modules, look at the system architecture for interactive and autonomous devices including 3D-sensing inputs, and then review the RealSense technology including its hardware modules and software libraries, followed by a discussion of various application areas. Finally, we summarize with a look into the future.

II. HUMAN 3D-SENSING SYSTEM

While we use all our senses to understand and interact with the world, here we focus on our visual and vestibular systems. These two perceptual systems function collaboratively to enable us to create a real-time 3D spatial map and use this information to understand and navigate in the 3D world around us. We briefly review the key aspects of our 3D-sensing and perceptual system, and refer interested readers to more detailed discussion [1, 2]. An understanding of our biological sensory systems and perceptual processes gives us important insights toward developing technologies and systems that can mimic human-like sensing and navigation capabilities.

First, let us consider the human visual system, which is depicted in Fig. 1. The human eye is a sophisticated visual imaging and transduction device. Light from the physical world enters through the pupil, and is focused on the retina at the back of the eye by the cornea and the lens combination, forming a projected two-dimensional (2D) image of the 3D world. Our visual sensing system comprises a binocular imaging scheme, where the two eyes form two distinct images of the same 3D scene in the physical world, as they image and capture the scene from two different locations. Due to this geometric construction, a point in the 3D space in front of us projects to two distinct locations on the retina of the left and the right eyes. The physical separation of the corresponding points on the retina is called the binocular disparity, which is inversely proportional to the distance of the physical point from the viewer. The human visual information processing system compares the two projected images formed by the left and the right eyes to discern the binocular disparity map corresponding to the 3D scene. This mechanism, termed stereopsis, along with a number of additional important monocular 3D cues and other sensory information, helps create a 3D spatial reconstruction of the visual world.

In addition, the human vestibular sensory system provides information on our movement, spatial orientation and balance, which are crucial to our abilities of real-time positional tracking and navigation in the 3D world. As depicted in Fig. 2, the sensors of the vestibular system include a set of three semi-circular canals in each of the inner ears, which indicate rotational movement information detected by the motion of the fluidic material inside the canals caused by our movements in the 3D space. In addition, a set of otoliths, the three oval-shaped structures in each of the inner ears, measure the linear accelerations resulting from our movements. The visual and vestibular systems work in synchronicity, providing consistent spatial and motion cues, thereby enabling us to create and update the 3D map of the environment around us in real-time.
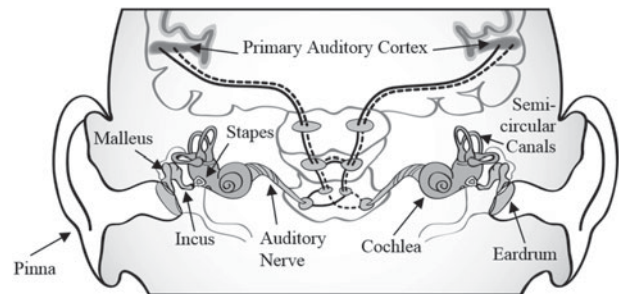

Fig. 2. The semi-circular canals in the inner ear, along with otoliths, form the sensors of the human vestibular sensing system. While the ears and the auditory cortex help us sense and understand sound, the vestibular system provides important cues on movement, orientation, and balance [1].
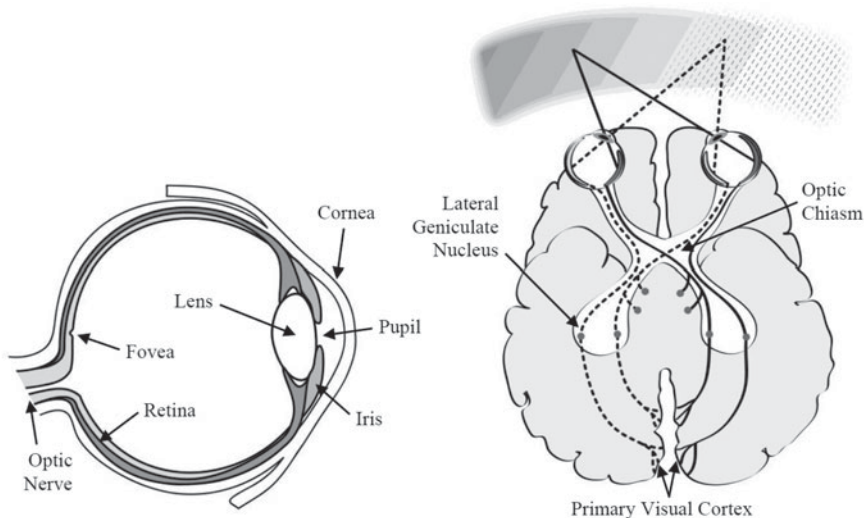

Fig. 1. Human visual system: left figure shows the construction of the human eye, right figure shows the binocular 3D imaging scheme. The depth or distance of the objects in the scene is discerned from the binocular disparity along with other visual cues such as motion parallax, occlusion, focus, etc. [1].

The 3D visual and vestibular systems, including the sensor modules and corresponding processing functions in the cerebral cortex, are not unique to humans. In fact such systems, albeit with varying degrees of acuity and capabilities, are common in nature among mammals who acquired them via millions of years of evolutionary process. As we describe below, development and incorporation of technologies inspired by these mechanisms enable us to architect and build intelligent systems with interactive and autonomous functionalities.

III. 3D-SENSING TECHNOLOGIES AND SYSTEMS

In general, interactive and autonomous systems consist of input technologies for receiving information from the environment or instructions from the users, computing technologies to execute processing functions according to the inputs, and actuator technologies to perform actions as the output of the processing. The block diagram, shown in Fig. 3 depicts the generic functional modules and flow of an interactive device or a system. The interactions between the device and the environment or the users are orchestrated by the interfaces, namely the inputs and the actions modules shown in the beginning and at the end, respectively.

The inputs module consists of sensors that transform the physical input stimuli from the environment or the users into electrical signals, while the action module provides
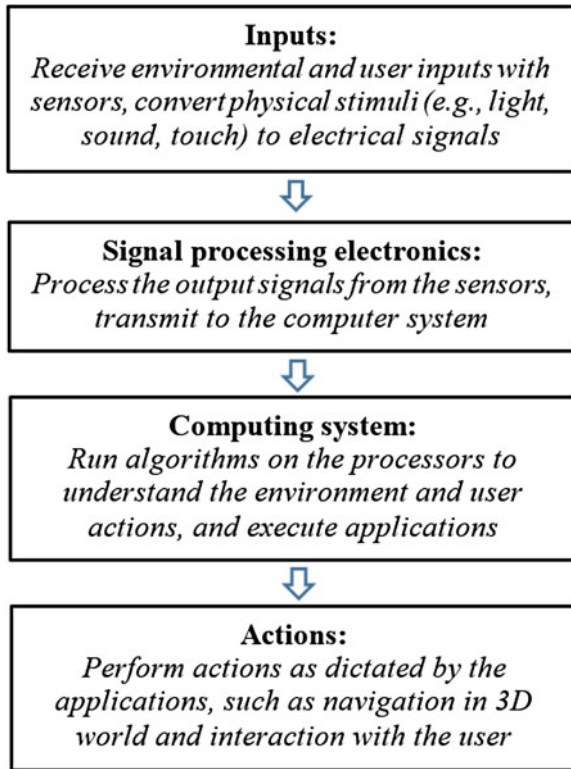

Inputs:
Receive environmental and user inputs with sensors, convert physical stimuli (e.g., light, sound, touch) to electrical signals

⇩

Signal processing electronics:
Process the output signals from the sensors, transmit to the computer system

⇩

Computing system:
Run algorithms on the processors to understand the environment and user actions, and execute applications

⇩

Actions:
Perform actions as dictated by the applications, such as navigation in 3D world and interaction with the user

Fig. 3. Functional block diagram of an interactive system. The inputs and the actions modules orchestrate the interactions between the system and the world or user, while the signal processing and computing functions facilitate these interactions.

the responses back to the user such as a screen displaying visual information, or a speaker producing audio output, or robot-performing physical actions such as navigating in the 3D world. The blocks in between perform the necessary processing and computing functions to facilitate these interactions.

We will first consider the imaging technologies for interfaces and interactions based on visual sensing. Cameras are now ubiquitous part of many devices and systems, as numerous applications based on capturing and consuming visual images have become part of our daily lives. However, the traditional cameras that are embedded in typical electronic devices are designed to capture 2D images of the 3D scene projected onto a single image sensor by the optics of the imaging system. This process can be represented by a matrix formalism as shown in the following equation, where the 3D points in the world are mapped onto a corresponding array of 2D points on the imaging device with a combination of necessary transformation matrices consisting of rotation and translation of coordinate systems as well as a perspective projection matrix.

$$
\begin{pmatrix} 2\mathrm{D} \\ \mathrm{point} \end{pmatrix} = \begin{pmatrix} \text{Camera to} \\ \text{pixel coordinate} \\ \text{transformation matrix} \end{pmatrix} \begin{pmatrix} \text{Perspective} \\ \text{projection matrix} \end{pmatrix}
$$

$$
\times \begin{pmatrix} \text{World to} \\ \text{camera coordinate} \\ \text{transformation matrix} \end{pmatrix} \begin{pmatrix} 3\mathrm{D} \\ \mathrm{point} \end{pmatrix}.
$$

As a result of this projection, the original 3D information cannot be generally recovered from the 2D images that are captured, since the captured 2D images preserve only a partial information about the original 3D space. Reconstruction of 3D surfaces from single intensity images is a widely researched subject, and continues to make significant progress [3]. However, implementation of real-time 3D interaction applications based on the single 2D image-sensing devices remains limited in scope and computationally intensive.

In contrast, 3D-imaging techniques, typically consisting of acquisition of a pair of color and depth images corresponding to the 3D scene, are designed to capture the 3D visual information. There has been significant progress in the areas of 3D visual sensing technologies in the recent years, resulting in small form-factor imaging systems that are able to capture both color and 3D spatial information in real-time with low-power consumption. While there are many ways for real-time 3D visual sensing, the prevalent methods are stereo-3D imaging, structured- or coded-light projection systems, and time-of-flight range imaging techniques. We describe the overviews of these techniques below. Chapters 5–7 in [1] provide in-depth details of the working principles of various 3D-sensing technologies.

Stereo-imaging-based 3D computer vision techniques attempt to mimic the human-visual system, in which two calibrated imaging devices, laterally displaced from each other, capture synchronized images of the scene. The depth
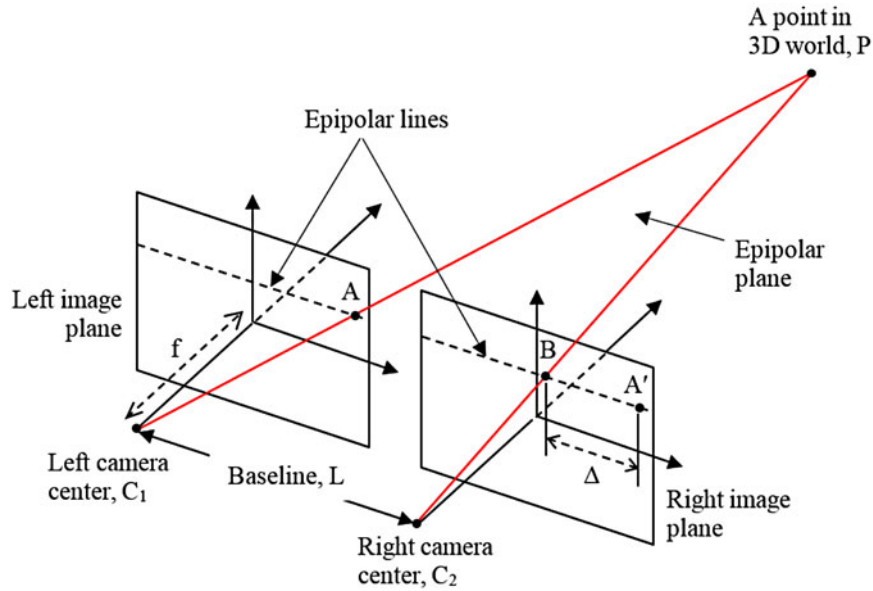
Fig. 4. Basics of the stereo-3D imaging method, illustrated with the simple case of parallel and calibrated camera pair with optical centers at C_1 and C_2, respectively, separated by the baseline distance of L. The point, P, in the 3D world is imaged at points A and B on the left and the right cameras, respectively. A′ on the right image plane corresponds to the point A on the left image plane. The distance between B and A′ on the epipolar line is called the binocular disparity, Δ, which can be shown to be inversely proportional to the distance of the point P from the baseline [1].

for the points in the 3D space mapped to the corresponding image pixels is extracted from the binocular disparities. The basic principles behind this technique are illustrated in Fig. 4, where C_1 and C_2 are the two camera centers with focal length f, forming images of a point in the 3D world, P, at positions A and B in their respective image planes. In this simple case, where the cameras are parallel and calibrated, it can be shown that the distance of the object, perpendicular to the baseline connecting the two camera centers, is inversely proportional to the binocular disparity: depth $= f \times L / \Delta$. Algorithms for determining binocular disparity and depth information from stereo images have been widely researched and further advances continue to be made.

In the case of structured-light-based 3D-sensing methods, a patterned or "structured" beam of light, typically infrared (IR), is projected onto the object or scene of interest. The image of the light pattern deformed due to the shape of the object or scene is then captured using an image sensor. Finally, the depth map and 3D geometric shape of the object or scene are determined using this distortion of the projected optical pattern. This is conceptually illustrated in Fig. 5 [4]. Further advances have been made on these techniques, such as using time-multiplexed binary code patterns to assign a unique digital code to each point indicative of its location in the 3D space, as described in Chapter 5 of [1].

The time-of-flight 3D imaging method measures the distance of the object points, hence the depth map, by illuminating an object or scene with a beam of modulated IR light and determining the time it takes for the light to travel back to an imaging device after being reflected from the object or scene, typically using a phase-shift measurement technique [1]. The system typically comprises a full-field range imaging capability, including amplitude-modulated illumination source and an image sensor array. Figure 6 illustrates
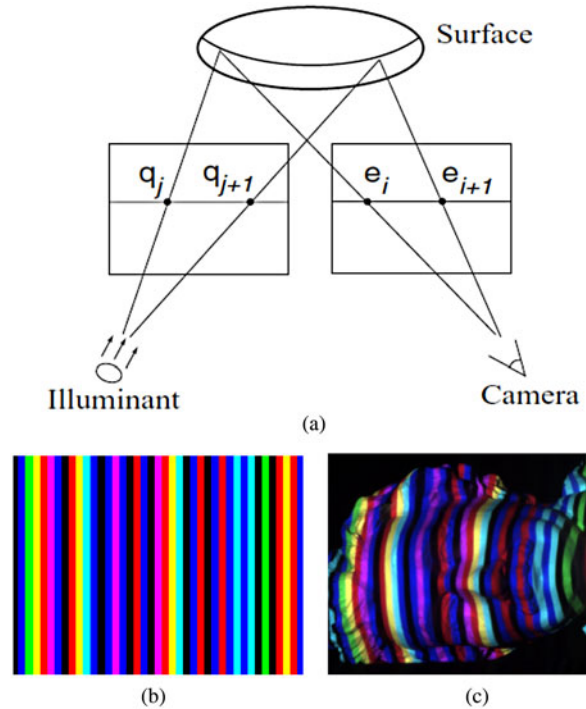

Fig. 5. Principles of a projected structured-light 3D image capture method [4]. (a) An illumination pattern is projected onto the scene and the reflected image is captured by a camera. The depth of a point is determined from the relative displacement of it in the pattern and the image. (b) An illustrative example of a projected stripe pattern. In practical applications, typically IR light is used with more complex patterns. (c) The captured image of the stripe pattern reflected from the 3D object.

the method for converting the phase shifts of the reflected optical signal to the distance of the point. The reflected signal, shown in the dashed curve, is phase-shifted by ϕ relative to the original emitted signal. It is also attenuated in
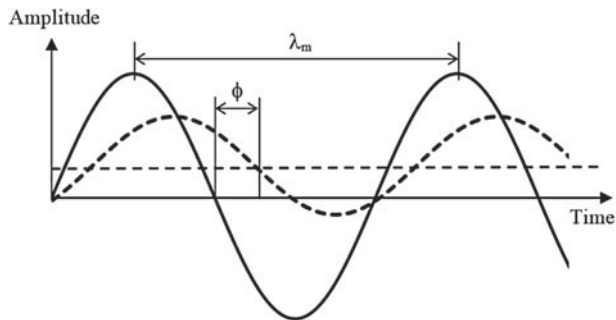
Fig. 6. Principles of 3D imaging using the time-of-flight-based range measurement technique [1]. The solid sinusoidal curve is the amplitude-modulated IR light that is emitted onto the scene by a source, and the dashed curve is the reflected signal that is detected by an imaging device. Note that the reflected signal is attenuated and phase-shifted by an angle ϕ relative to the emitted signal, and includes a background signal that is assumed to be constant. The distance or the depth map is determined using the phase shift and the modulation wavelength.

Fig. 7. Intel® RealSense camera modules. The top figure shows the F200 version based on coded-light 3D-imaging technique, whereas the bottom figure shows the R200 version based on stereo-3D imaging technique. The imaging processors consist of power-efficient hardware for 3D computation and processing.

strength and the detector picks up some background signal as well, which is assumed to be constant. With this configuration, it can be shown that the distance of the object that reflected the signal, $d = (\lambda_m/2) \times (\phi/2\pi)$, where λ_m is the modulation wavelength of the optical signal.

Besides the advances in 3D visual sensing technologies as narrated above, developments in inertial measurement units (IMU) are allowing incorporation of motion sensing capabilities into devices and systems. With accelerometers, gyroscopes, and often magnetometers integrated into small form-factor devices, systems are able to use positional tracking data from the IMUs for navigation purposes. A new and exciting method for real-time 3D tracking involves visual-inertial odometry, which combines the features and motions measured by the visual sensors as well as the IMUs for accurate localization, tracking, and mapping [5].

IV. INTEL® REALSENSE TECHNOLOGY

In this section, we describe the RealSense technologies and a series of products based on these technologies that have been introduced to the market, incorporating real-time 3D-sensing and interaction capabilities into various classes of devices and systems. The RealSense technology includes hardware sensor modules for capturing the 3D environment via real-time color and depth (RGB-D) imaging and 3D motion sensing, and a set of middleware libraries included in software developments kits to enable interactive applications and usages that utilize the 3D information (www.intel.com/realsense; www.intel.com/realsense/developer).

As examples, Fig. 7 shows two of the RealSense modules that have been integrated in various interactive and autonomous devices and systems. The module shown in the top figure is the RealSense F200 device, which is based on the coded-light 3D-sensing technology. As illustrated in the figure, this module consists of an IR laser and a microelectromechanical systems projector to illuminate the
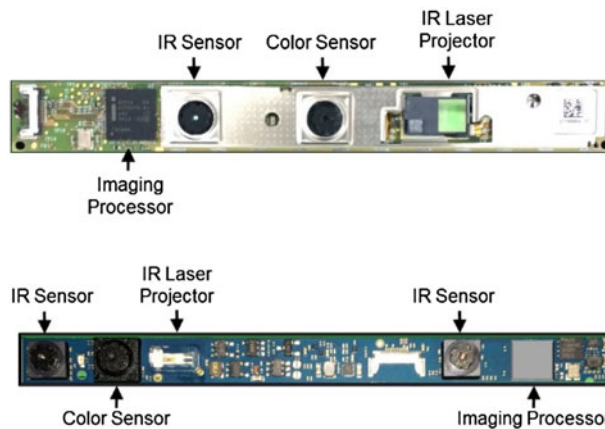
environment in front of it with specific binary IR patterns. An IR image sensor on the module rapidly captures the images of these patterns that are reflected from the 3D scene. At the same time, a color camera which is also part of the module captures high-resolution RGB images. A custom-built special-purpose processor on the module runs algorithms designed to compute the depth maps in real-time from the captured binary codes, which are synchronized and calibrated with the corresponding color images. The pairs of color and depth images are made available to the middleware layers and applications running on the computing systems via a single USB 3.0 interface, which also provides the power to the module.

The bottom figure shows the RealSense R200 module, which is based on IR-assisted and hardware-accelerated stereo-3D imaging technology. The illumination subsystem on the module projects a texture of IR light onto the 3D objects and scene in front of the camera system. The two IR image sensors separated by a baseline distance capture real-time images of the environment. An onboard imaging processor hardcoded with power-efficient algorithms runs rectification and stereo-correlation algorithms to compute binocular disparities and the corresponding depth images in real time. At the same time, the color image sensor captures high-definition RGB images, which are synchronized with the corresponding depth maps. Similar to F200 module, a single USB 3.0 interface provides power to the sensor module as well as transmits the data.

Both devices are <4 mm in thickness, enabling integration into a wide array of computing devices and systems. As narrated above, the sensor modules capture pairs of color and depth images, also referred to as RGB-D images, in real time. Figure 8 shows a pair of RGB-D images of a scene, recorded with a RealSense camera. Every pixel on the image shown in the left indicates the color value associated with the corresponding point in the 3D space, and that on the image on the right indicates the depth of the point from the sensor module. With these pair of images, the 3D world
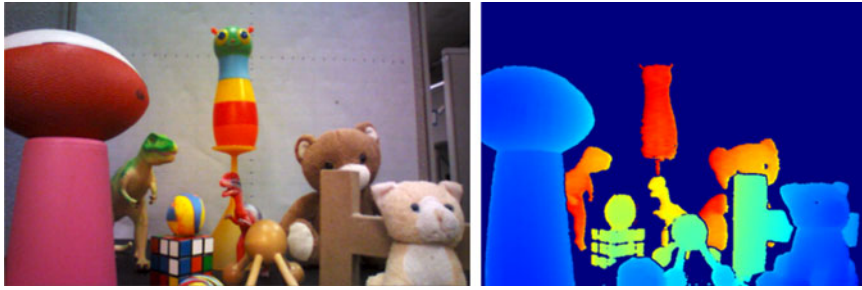
Fig. 8. A pair of RGB-D images captured with Intel® RealSense camera. The left figure shows a color image, while the right figure shows the corresponding pseudo-colored depth image where the nearer points are shown in bluer colors and farther points are shown in redder colors. The background objects that are further away from the range of the depth sensor are shown in dark blue.

Fig. 9. Examples of 3D computer vision middleware libraries included in the RealSense software development kit. Top left: 3D hand skeleton tracking; top right: face detection and tracking; bottom left: 3D background segmentation; bottom right: 3D scanning and reconstruction.

is captured by the RealSense devices in real-time, complete with both color and 3D coordinate information for the points, enabling development of interactive and intelligent systems. An additional module, which has been recently introduced but not shown in Fig. 7 is RealSense ZR300, which incorporates a wide field-of-view fish-eye camera and an IMU in addition to an R200 RGB-D sensor, all integrated onto a common stiffener, synced, and calibrated for accurate visual-inertial 3D-tracking applications as will be described later.

Besides the 3D-sensing hardware modules, the RealSense software development kit includes a number of middleware libraries and application programming interfaces to enable the development of a new class of interactive applications. Figure 9 illustrates a few of the 3D computer vision middleware technologies, including a hand skeleton tracking library, a face detection and tracking library, a background segmentation library, and a 3D reconstruction library, among many other capabilities.

V. SYSTEMS AND APPLICATIONS

RealSense devices have been integrated in a wide array of computing devices available from a number of system makers, including interactive computers, mobile devices, autonomous robots, UAVs, virtual dressing mirrors, augmented reality helmets, and among many other emerging applications. In this section, we highlight a number of different types of devices and systems that are already available in the market and demonstrate a wide range of applications that are enabled by real-time 3D-sensing technologies.

A) Interactive computing devices

A number of computing devices have incorporated RealSense cameras and are commercially available. Figure 10 shows a couple of examples, including a state-of-the-art all-in-one desktop computer with a curved display and front-facing RealSense F200 camera, and a 2-in-1 tablet

Fig. 10. Examples of commercially available computers with embedded RealSense technologies. Left: an interactive all-in-one desktop computer, right: a 2-in-1 laptop/tablet device.

with rear-facing RealSense Roo camera. Numerous applications have been developed based on the real-time RGB-D imaging and 3D computer vision middleware libraries, including 3D interactive games, login applications using facial recognition, video conferencing using virtual green-screen effects, and 3D scanning of humans and objects, to highlight just a few (www.intel.com/realsense/apps). We have also demonstrated and reported on a smartphone device with integrated 3D-sensing technology based on RealSense [5]. Figure 11 shows an example of dense 3D reconstruction of large spaces with this device.

B) Autonomous robots and UAVs

Among the most exciting areas of applications for real-time 3D-sensing and spatial tracking technologies are robots and drones that can sense and understand the environment around them, navigate autonomously, and interact

naturally with humans. Robots have already had a major positive impact to the world economy by automating the industrial manufacturing and assembly lines. This has significantly increased the production throughput across numerous sectors, spanning semiconductor chips to consumer electronics devices to automotive assembly processes to food production and processing, to name just a few among numerous examples. In this section, we highlight the capabilities of the family of RealSense technologies that are enabling a new generation of autonomous and intelligent machines.

As described in Section II, we the humans use our visual-vestibular spatial sensing and 3D-tracking capabilities for mapping and navigating in the 3D world. The RealSense technology attempts to endow devices and machines with sensing and perception capabilities inspired by such biological systems. For example, the RealSense ZR300 module comprises a depth-sensing module, a fish-eye camera, an inertial motion capture device, and a time-stamping circuitry to synchronize the inputs from all the sensors for multi-sensory tracking and navigation applications. We have implemented a real-time visual-inertial odometry solution based on this platform, which is shown in Fig. 12. The image in the middle shows the view from the fish-eye camera, while the figure on the left shows the 2D map view as the device navigates through a large 3D space. This technology allows devices and systems to map a 3D environment, localize within the map, and autonomously navigate in the 3D space. As shown in the right image of Fig. 12, such implementations can work well in a relatively large space,
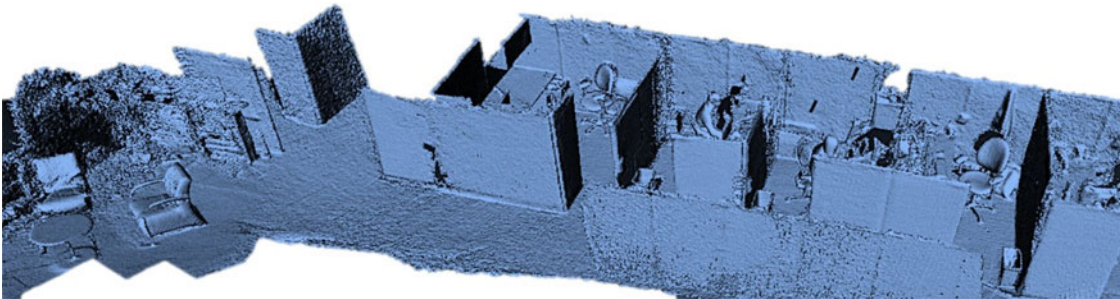

Fig. 11. Dense 3D reconstruction of an office environment captured with a mobile device incorporating RealSense technology. The real-time depth imaging with high-density point cloud allows rapid reconstruction of 3D spaces, objects, and humans.
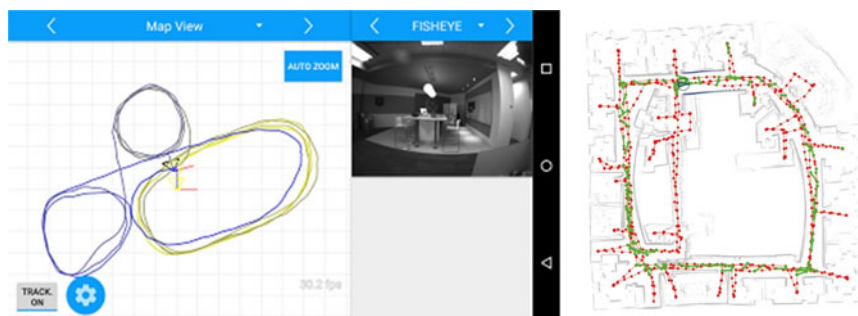

Fig. 12. Real-time 3D spatial tracking with six degrees of freedom using visual-inertial odometry. The image in the middle shows the view from the fish-eye camera, the image on the left shows the 2D view map traced while navigating within a 3D space. The figure on the right shows large-scale 3D mapping and navigation spanning an entire floor of a office building.

Fig. 13. Examples of autonomous robots equipped with RealSense technology. Left, a hotel butler robot from Savioke; middle, Segway personal transporter robot from Ninebot; right, a personal assistant home robot from Asus.

Fig. 14. The left image shows the Yuneec Typhoon H drone with integrated RealSense device as demonstrated in CES 2016. The right image shows a demonstration of real-time automatic collision avoidance as the drone follows a person biking through trees.

which makes it suitable for autonomous robotic navigation applications.

A number of autonomous robots with various application areas have been introduced which incorporate RealSense devices. As shown in Fig. 13, some of the examples include a hotel butler robot from Savioke that autonomously navigates within a hotel to deliver items to the guests [6], a multipurpose Segway personal transportation robot unveiled by Ninebot in the 2016 Consumer Electronics Show which can interact with the users and the environment [7], and an intelligent home assistant robot demonstrated by Asus in the 2016 Computex show [8].

UAVs, also referred to as drones, are a fast-growing market with applications ranging from consumer and professional photo and videography to commercial inspections to automated deliveries in the future. We have added real-time 3D-sensing capability to drones by integrating RealSense modules. With real-time depth-imaging technology provided by RealSense, we have implemented a collision avoidance solution on the drone, enabling it to safely and automatically fly around trees and other objects without hitting into them. Figure 14 shows a drone from Yuneec with a RealSense device onboard [9] and the image from a demonstration of automatic collision avoidance function, while the drone follows a biker on a trail.

C) Virtual and augmented reality devices

Now, we will focus on the application of real-time 3D-sensing and visual-inertial tracking technologies in a new class of virtual and augmented reality devices that enable immersive and interactive mixed-reality usages. The development of virtual and augmented reality devices and applications has picked up significant pace around the world in the recent years. Let us first look at the key definitions of the current systems based on their usages and applications. A virtual-reality device places the user in a virtual environment, generating sensory stimuli (visual, vestibular, auditory, haptic, etc.) that provide sensation of presence and immersion. On the other hand, an augmented-reality device places virtual objects in the real-world while providing sensory cues to the user that are consistent between the physical and augmented elements. While a virtual-reality device immerses the user within a simulated environment, it also removes the user from the surrounding real world. In contrast, a mixed-reality device can blend real-world elements within the virtual environment.

We have demonstrated interactive mixed-reality applications based on embedded RealSense and visual-inertial spatial and motion tracking algorithms [10]. As an example, the picture of a prototype head-mounted device is shown in the left image of Fig. 15. The visual-inertial tracking and real-time 3D-sensing capabilities allow the device to map the 3D environment around it, localize and track the positional information with six degrees of freedom. This enables immersive navigation in the virtual space without requiring external tracking systems. As shown in the right image in Fig. 15, the RealSense 3D-imaging technology also enables integrating the user's real hands into the simulated environment for direct interactions and manipulations of the virtual objects in the 3D space. Figure 15 also shows further mixed-reality capabilities, as a person standing in front of the user is brought into the virtual world as viewed with the virtual-reality headset. Besides enabling natural and immersive interactions, this technology also allows the user to avoid colliding into objects in the physical world while moving
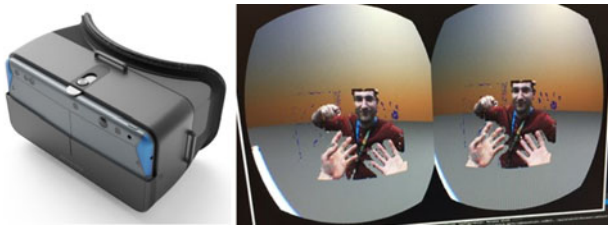
Fig. 15. Left image shows an interactive mixed-reality device incorporating RealSense and visual-inertial spatial motion tracking technology. The image on the right shows an example of mixed-reality capability of the device, where the 3D images of the user's hands as well as a person standing in front of the user are brought into the virtual world. This capability is also used to allow the user to avoid colliding into physical objects.

Fig. 16. Augmentation of the real physical world with virtually rendered 3D objects using a device with embedded RealSense module. Here a digitally rendered car is shown racing on a real kitchen table and colliding into a physical bowl, with realistic physical effects such as collision with real objects, correct occlusion, shadows, etc.

about in the virtual world. Finally, Fig. 16 demonstrates an example of augmenting the real-world with virtually created objects with correct physical interactions, such as collisions, occlusions, shadows, etc., where a virtual car races on a real kitchen table.

D) Emerging applications

Besides the areas discussed above, there are numerous other interactive and intelligent systems and applications that are being enabled by the real-time 3D-sensing and RGB-D imaging technologies incorporated in RealSense devices. Examples include 3D body scanning for fitness tracking, virtual clothing appliances, interactive gaming peripherals, sporting, and entertainment applications.

VI. SUMMARY

In this paper, we have reviewed the recent developments in the field of 3D-sensing technologies, systems, and applications. As an example of a commercially deployed platform, we have described the 3D-sensing technologies with real-time RGB-D imaging and 3D spatial tracking capabilities incorporated in the Intel® RealSense cameras, and the associated 3D computer vision and spatial-understanding algorithms and middleware libraries. We have reviewed a number of applications of these technologies in intelligent

systems, including a new class of interactive computing devices, autonomous machines such as robots and UAVs, and immersive mixed-reality devices that blend real-world objects into the virtual world and enable natural interactions.

It has been a dream for many of us to add human-like sensing, understanding, and navigation capabilities to devices and machines. The rapid developments in the field of perceptual computing, spanning advanced sensors, computing hardware, intelligent algorithms, autonomous systems, and applications are bringing us closer to this dream than ever!

ACKNOWLEDGEMENTS

REFERENCES

[1] Bhowmik, A.K. (Ed.): Interactive Displays: Natural Human-Interface Technologies, Wiley & Sons, 2014. Available: http://www.wiley.com/WileyCDA/WileyTitle/productCd-1118631374.html

[2] Goldstein, E.B.: Sensation and Perception. Cengage Learning, 2013. Available: http://www.cengage.com/search/productOverview.do?N=14+82&Ntk=P_EPI&Ntt=9409927561299139231504084516447852508&Ntx=mode%2Bmatchallpartial

[3] Szeliski, R.: Computer Vision, Springer, 2013. Available: http://www.springer.com/us/book/9781848829343

[4] Zhang, L.; Curless, B.; Seitz, S.M.: Rapid shape acquisition using color structured light and multi-pass dynamic programming, in 3DPVT, Padova, 2002.

[5] Bhowmik, A.K. et al.: Immersive applications based on depth-imaging and 3D-sensing technology, in *SID Symp. Digest of Technical Papers I1.3*, San Jose, 2015.

[6] Piltch, A.: Robotic butler answers your room service call. Available: www.tomsguide.com/us/savioke-relay-robotic-butler,news-21488.html. Retrieved on 24 July 2016.

[7] Plummer, Q.: segway's hoverboard robot uses Intel Realsense to find its way around. Available: www.techtimes.com/articles/122451/20160107/segways-hoverboard-robot-uses-intel-realsense-to-find-its-way-around.htm. Retrieved on 24 July2016.

[8] Wang, Y.: Asus unveils its first home robot product at COMPUTEX. Available: en.ctimes.com.tw/DispNews.asp?O=HK05UCJVQ66SAA00N5. Retrieved on 24 July 2016.

[9] Goldman, J.: Yuneec Typhoon H drone gets new obstacle-avoiding powers from Intel. Available: www.cnet.com/news/yuneec-typhoon-h-drone-gets-new-obstacle-avoiding-powers-from-intel. Retrieved on 24 July 2016.

[10] Pleasant, R.: Hands on: virtual reality with Intel RealSense. Available: siliconangle.com/blog/2016/03/14/hands-on-virtual-reality-with-intel-realsense-at-gdc-2016. Retrieved on 24 July 2016.

Dr. Achintya K. Bhowmik is vice president and general manager of the Perceptual Computing Group at Intel Corporation. He leads the development and deployment of advanced computing solutions based on natural sensing and intelligence,

branded as Intel® RealSense Technology. His responsibilities include creating and growing new businesses in the areas of interactive computing systems, immersive virtual reality devices, autonomous robots and unmanned aviation systems. Previously, he served as the chief of staff of the personal computing group, Intel's largest business unit. Prior to that, he led the development of advanced video and display processing technologies for Intel's computing products. His prior work includes liquid-crystal-on-silicon microdisplay technology and integrated electro-optical devices. As an adjunct and guest professor, Dr. Bhowmik has advised graduate research and taught human-computer interactions, computer vision and display technologies at the Liquid Crystal Institute of the Kent State University, Stanford University, University of California, Berkeley, Kyung Hee University, Seoul, and the Indian Institute of Technology, Gandhinagar. He has over 100 publications, including two books titled "Interactive Displays: Natural Human-Interface Technologies" and "Mobile Displays: Technology & Applications", and over 100 granted and pending patents. Dr. Bhowmik was elected Fellow of the Society for Information Display (SID) in 2016. He received the Industrial Distinguished Leader Award from the Asia-Pacific Signal and Information Processing Association (APSIPA) in 2016. He serves on the executive committee and board of directors for SID, and is an associate editor for the Journal of SID. He is on the board of directors for OpenCV, the organization behind the open source computer vision library.