New Opportunities for Computer Vision–Based Assistive Technology Systems for the Visually Impaired

Juan R. Terven and Joaquín Salas, *Instituto Politécnico Nacional, Mexico*
Bogdan Raducanu, *Computer Vision Center, Spain*

Computing advances and increased smartphone use gives technology system designers greater flexibility in exploiting computer vision to support visually impaired users. Understanding these users' needs will certainly provide insight for the development of improved usability of computing devices.

Computer vision as a substitute for human vision constitutes a powerful tool for developing assistive technologies for the visually impaired. Applications based on computer vision enhance these persons' mobility and orientation as well as object recognition and access to printed material. More recent systems even attempt to support social interaction.

Such assistive technology is critical to a sizable percentage of the global population. According to the World Health Organization, in 2013, approximately 39 million people were blind, and another 246 million had some kind of moderate or severe visual impairment (www.who.int/mediacentre/factsheets/fs282). The same survey also showed trends in the causes of visual impairment. In developed countries, the major causes of blindness were degenerative diseases, but in underdeveloped countries, vision loss stemmed from mostly treatable diseases, such as cataracts. In these countries, visual loss tended to accompany impairment in other senses, including hearing or touch.

Understanding the significance of these trends should be integral to the development of any assistive technology for the visually impaired. Developers tend to miss solutions that consider the human side of the user experience, such as the need to operate as inconspicuously as possible in a society oriented to sighted individuals. The design must be user centered so that the user's needs are taken into account from conceptualization up to final prototype validation. Technology-based solutions are important options, but cultivating human capacities is also vital. For example, many studies have demonstrated the value of acknowledging human ingenuity in problem solving, such as recognizing visually impaired individuals' ability to navigate using echolocation.[1]

Computer vision systems are ripe for fine-tuning the intersection between human potential and advanced technology. Generally, assistive technologies for people with visual impairment have been based on ultrasound, infrared, or laser sensors, and for some time developers have wanted to change that technology base to artificial vision to achieve the same goal. Until a decade ago, limited computing power and a paucity of reliable algorithms have blocked progress in this direction. Now, however, computing has

 Published by the IEEE Computer Society

advanced to the point that reliable vision algorithms can execute in real time on embedded computers. In addition, the convenience and pervasiveness of smartphones have led to the development of vision-based technologies that help visually impaired users participate in a variety of daily activities that others take for granted.

Unlike most systems based solely on ultrasound, infrared, or laser technologies, computer vision systems use advanced scene-understanding algorithms to reconstruct the environment around a visually impaired user. The tradeoff is increased processing complexity, but so far application requirements are not exhausting the core processor's ability to meet them. Indeed, the integration of heterogeneous computing systems and digital cameras in smartphones and tablets has stimulated exciting new applications—from providing greater mobility to making social interaction more enjoyable.

To understand the impact of these applications, we surveyed the literature on computer vision applications, which complements our work at the Instituto Politécnico Nacional (IPN) to develop prototype computer vision systems for object detection and enhanced social interaction. Our survey looked at developments in mobility, orientation, object detection, print access, and social interaction.

MOBILITY

Independent travel in unknown environments is a daily challenge for visually impaired people. To detect and avoid obstacles most individuals use a white cane, but it fails to protect against head-level obstacles. Some individuals rely on a guide dog, which can help the individual navigate more effectively, but the costs of acquiring and training a guide dog might be prohibitive. Clearly, these limitations motivate the development of assistive technology for increased mobility.

Computer vision–based electronic travel aids

Electronic travel aids (ETAs) have been available since the 1960s, but an inadequate interface and lack of usability have made them unsuitable for visually impaired users. ETAs require either an acoustic or haptic interface, which is designed to offer a sensory substitution for vision, such as vibrations. However, humans see orders of magnitude more than they hear or touch,[2] so interfaces based on these other senses cannot fully substitute for vision. In addition, ETAs must be able to provide adequate feedback from the wealth of perceived information about the user's environment. Feedback must be fast and not interfere with hearing and touch. Another constraint is size. The system must be unobtrusive (usually embedded), lightweight, and comfortable.

Computer vision is helping researchers meet these requirements, although few systems are commercially available. Table 1 lists the computer vision–based ETAs that we surveyed—one a commercial system and the rest more recent laboratory prototypes, including the two we have developed at IPN.

Of the few commercial computer vision–based ETAs developed since the 1990s, most rely on ultrasonic signals. Among these is the vOICe system,[3] which transforms an image into a sound and transmits the sound to the user's headphones, which the user learns to recognize through training. The development of commercial ETAs most likely stalled because providing the required computing power was technically infeasible or too expensive. The emergence of multimedia processors capable of real-time image processing should rejuvenate commercial innovation.

Computing has advanced to the point that reliable vision algorithms can execute in real time on embedded computers, and smartphones have led to vision-based technologies that help visually impaired users participate in a variety of daily activities.

A growing research community dedicated to computer vision is creating laboratory prototypes, most of which use stereo vision to generate occupancy maps as part of detecting obstacles and determining their distance from the user. Examples[4] include Virtual Acoustic Space from the Instituto de Astrofísica de Canarias, Spain; the Electron-Neural Vision System from the University of Wollongong, Australia; the Tactile Vision System from the University of Arizona; and Tyflos from Wright State University.

Stereo vision's main drawback was computational complexity, which lessened with the appearance of low-cost depth cameras such as Microsoft's Kinect. These cameras directly input a depth map to the processor, thereby reducing the required calculations. A disadvantage is that these cameras tend to function properly only indoors because the infrared component of direct sunlight affects their operation. Nonetheless, in the past few years, several prototypes have incorporated these kinds of sensors. Among these are KinDetect[5] from City College of New York and our Vibratory Belt system.

Enhancing user mobility with the Virtual White Cane and Vibratory Belt

Regarding our research work at IPN, we developed the Vibratory Belt along with the Virtual White Cane[6] to enhance a visually impaired user's mobility. As Figure 1a shows, the

Table 1. Examples of computer vision–based electronic travel aids (ETAs) from oldest to most recent.

System	Functionality	Interface and components	Advantages (A) and disadvantages (D)	Study results
vOICe (1992)	Vision substitute; represents environment acoustically	Stereo acoustic: digital camera embedded in eyeglasses; headphones; portable computer	A: Portable, relatively small D: Blocks user's hearing; requires considerable training	Enough training to show promising results
Virtual Acoustic Space (1999)	Vision substitute; orientation by constructing an acoustic perception of the environment	Stereo acoustic: two digital cameras embedded in eyeglasses; headphones; portable computer	A: Portable, reduced form factor D: Blocks user's hearing; not tested in real environments	Six visually impaired and six sighted, showed >75% of object and distance detection
Electron-Neural Vision System (2005)	Obstacle detection by electric stimulation in both hands; each finger represents a zone in the frontal field of view	Two digital cameras; digital compass; laptop with GPS; gloves with stimulators in each finger	A: Real-time performance; does not block user's hearing D: Blocks use of both hands; does not detect ground- or head-level obstacles	All one-hour–trained blindfolded users could traverse a path avoiding obstacles
Tactile Vision System (2006)	Obstacle detection by vibrations across waistline through a vibrator belt	Belt with 14 vibrators; two digital cameras; laptop	A: Does not block hearing or hands D: Unable to distinguish between floor-level and hanging objects; needs testing with real users	No reports of experiments with visually impaired users
Tyflos (2008)	Obstacle detection by vibrations across chest through a vibrating vest	Vest with a 4 × 4 vibrator array; two digital cameras; laptop	A: Does not block hearing; detects obstacles at various height levels D: Needs more tests on real users	No experiments with visually impaired users
Kindetect (2012)	Person and obstacle detection by acoustic feedback	Acoustic depth sensor; computer	A: Easy to use; can detect head-level obstacles D: Blocks hearing; indoor operation only	Four blindfolded users detected obstacles along an indoor path
Vibratory Belt (2013)	Obstacle detection by vibrations across waistline through a vibratory belt	Belt with three motors; embedded computer; Kinect sensor	A: Easy to use; does not block hearing; detects head-level obstacles D: Generally, indoor operation only	18 blindfolded users performed comparisons between the white cane and the vibratory belt, resulting in similar travel time along an indoor path
Virtual White Cane (2013)	Obstacle detection by vibration of a Smartphone	Android Smartphone coupled with a laser pointer	A: Easy to use; does not block hearing; detects head-level obstacles D: Needs more tests on real users	18 blindfolded users detected obstacles along an indoor path

Virtual White Cane combines a laser pointer and smartphone to simulate a cane. The laser aligns with the camera to generate a baseline (t_x) and pan angle (α), as in Figure 1b. The camera captures the laser's reflection off the planar surface, which become input to a smartphone application that uses active triangulation to calculate the object's distance from the user. The application then uses the smartphone's vibration to alert the user to the object's proximity.

Figure 2a shows a user wearing the Vibratory Belt, which contains a Kinect camera connected to an embedded computer, an inertial measurement unit (IMU), and three small vibrating motors. The camera provides depth images, such as that in Figure 2b, and the computer calculates the distance of the closest obstacles in three positions in front of the user, as in Figure 2c. To separate the floor from the objects, the IMU calculates the camera's orientation.

Both the Virtual White Cane and the Vibratory Belt rely on a haptic interface to provide feedback to the user. To explore how the user perceived the generated vibrations, we performed several stimuli versus sensation trials with multiple individuals. The trials enabled us to fine tune the vibrations to represent a particular distance. This is critical as perception connects the information coming from the diverse senses to the user's knowledge, memories, and expectations. In previous experiments,[13] a 50-m path was defined in an office-like environment. Then, blindfolded and non–visually impaired users traversed the path in one of four ways: normally, using only their hands, using the

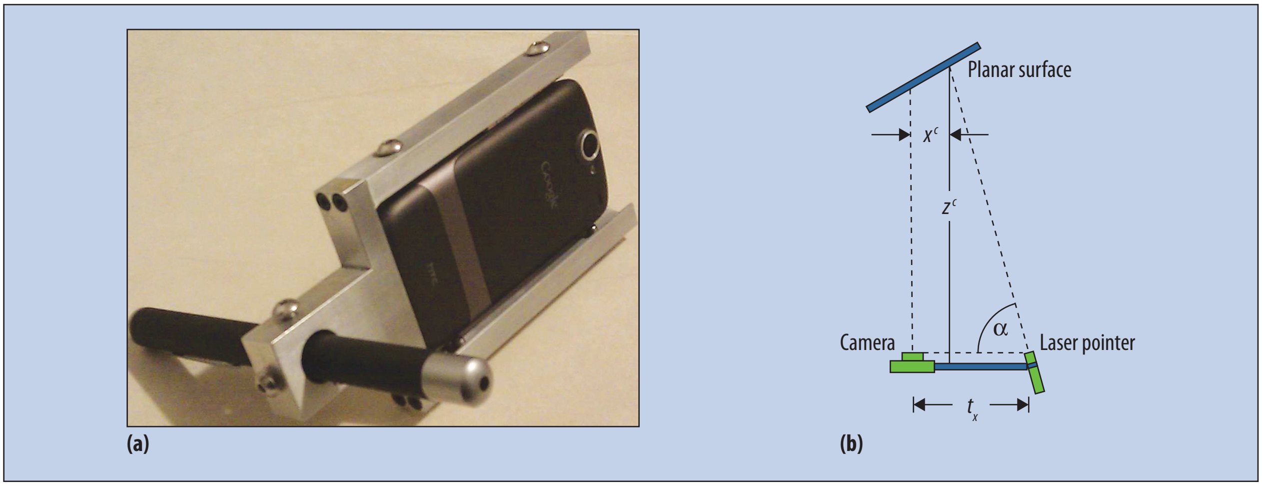

Figure 1. Virtual White Cane. (a) The user carries an Android Smartphone coupled with a laser pointer. The device simulates a white cane vibrating in the presence of obstacles where the user is pointing to. The metallic structure couples the laser pointer and the user's smartphone. (b) As the laser points to an object near the user, a smartphone application uses active triangulation to measure the object's distance (z^C) and causes the smartphone to vibrate. As the object gets closer, the vibration intensity increases.

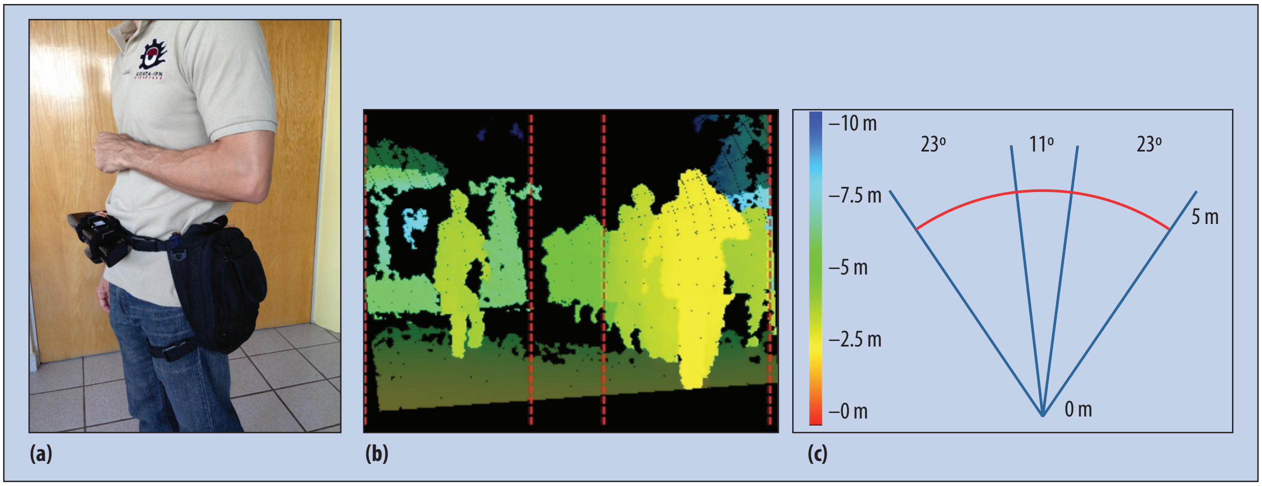

Figure 2. Vibratory Belt. (a) Complete prototype worn by the user, (b) depth map that the Kinect camera inputs to the embedded computer, and (c) obstacle map showing the position in degrees (horizontal axis) and the distance in meters (vertical axis).

Virtual White Cane, or using a regular white cane. The results show that the new instrument is more effective than hands-only, but is no better than the regular white cane.

ORIENTATION

Orientation is the capacity to know and track a current position with respect to the environment and to find a path to the desired destination. With Crosswatch,[7] a mobile application to help a visually impaired individual cross the street, the user finds crosswalks with stripes by pointing the smartphone's camera to the street. The system uses pattern recognition to analyze images of the street and produces an audible tone when it detects a crosswalk. The system works well as long as the crosswalk has stripes.

Another common orientation problem for visually impaired individuals is how to know their current position and reach another location. An algorithm for indoor orientation[8] detects partial door edges and corners and constructs a geometric model with four connected corners. Because the algorithm uses contours, it can detect open doors under various illuminations, deformations, and scales.

Some indoor orientation algorithms are based on the idea of using labels inside the building that cameras can easily detect. One algorithm[9] proposes using printed reflexive patterns to designate key locations, such as exits, elevators, and restrooms. Individuals wearing a regular smartphone camera could then identify the patterns. Another approach[10] proposes identifying key spots with labels that consist of figures with a defined shape and color, which computer vision algorithms could easily recognize.

MIT's Fifth Sense Project (http://people.csail.mit.edu/teller/misc/bocelli.html) aims to find a safe walking surface, detect collision hazards, and generally provide a rich set of orientation capabilities. It leverages surrounding cues, such as text signs and features, to provide the user with enough information to know current position and how to get to a desired destination, as well as to infer common locations such as a kiosk, concierge desk, elevator lobby, or water fountain. The system is an example of capitalizing on technology gains—in this case, decades of robotics technology knowledge—to create novel assistive technology devices.

Access to printed information can be considered the area where computer vision has been most successful due to advances in optical character recognition (OCR).

PRINT ACCESS

Reading anything from street signs to product information is a common activity that sighted people take for granted. In sharp contrast, only 10 percent of visually impaired children learn Braille and most documents are not available in this format (https://nfb.org/braille-campaign). Clearly, the development of reading devices is paramount.

Access to printed information can be considered the area where computer vision has been most successful due to advances in optical character recognition (OCR). Early OCR devices, such as the Arkenstone reader, were bulky and required the scanning of an entire page. Decades later, smartphone applications such as kReader (www.knfbreader.com/products-kreader-mobile.php) or Georgie (www.georgiephone.com) allow much finer-grained text selection, but a visually impaired user must still point the camera to frame the text, which can be frustrating.

To address this problem, Voiceye (http://voiceye.viewplus.com) features a 2.5-cm^2 code patch that can store up to two complete text pages. Users access the information by aiming a Smartphone with the Voiceye app to scan or photograph the material containing the small code. In South Korea, Voiceye codes are used by schools for the visually impaired, universities with special education, publishing companies, and large corporations such as LG. Researchers are working on detecting and recognizing other kinds of readable media, such as the LED and LCD displays prevalent in household appliances and irregular images that combine figures with text such as graphs, logos, and street signals. Solutions typically pair such readers with object detection and recognition algorithms.

OBJECT RECOGNITION

Object recognition is another promising practical application of computer vision. Unlike mobility applications where the system must recognize the presence and distance of any object (or obstacle), object recognition deals with the detection of known objects provided in a database. LookTel has developed two smartphone applications to recognize objects: Money Reader and Recognizer. Money Reader helps users differentiate paper currency by using a voice synthesizer to read the bill's value. Recognizer takes an image of an object and compares the image with an internal database that the user creates. Once it recognizes the object, the system reproduces the object description that the user previously recorded.

Trinetra,[11] a prototype system developed at Carnegie Mellon University, uses product barcodes to help users recognize supermarket objects, but the user must be able to direct the camera toward the barcode, which can be tedious. To overcome this limitation, an algorithm[12] helps locate the barcode by giving left or right indications until the user succeeds in focusing the camera on the code.

Another prototype application[13] recognizes supermarket objects that the user has put in a shopping list on the smartphone. As the user moves the camera through the aisles, the application detects the objects (using an internal database) and compares them with the ones on the list, notifying the user when it finds a match.

All these applications use object recognition as a way to provide visually impaired individuals with greater autonomy as they shop. Another application[14] takes a slightly different twist by enabling users to recognize bus numbers without relying on another person. The algorithm combines geometric computer vision with machine learning to achieve robustness against reflections, shadows, and partial occlusions.

SOCIAL INTERACTION

Much human interaction involves nonverbal actions, such as smiling, winking, frowning, and gesturing—all of which are inaccessible to a visually impaired person. This limitation can cause feelings of social exclusion and alienation.

iCARE Social Interaction, a project at Arizona State University, empowers visually impaired individuals during social encounters by providing visual information about those around them. The prototype system uses a camera embedded in eyeglasses that communicates with a smartphone.[15] Employing computer vision algorithms, iCARE detects the other person's position and transmits this information to the user through a vibrating belt. The system can also detect six basic emotions—happiness, sadness, surprise, anger, fear, and disgust—as well as a neutral emotional state. It then provides this information to the user through a glove with 14 small vibratory motors in the fingers.

Deciding on the degree and type of user feedback is a major design challenge because a large percentage of information about a social environment is impossible to transmit through acoustics or vibrations. Enhancing social interaction requires intelligent algorithms that can extract multiple features but are discerning enough to supply only the most important social cues in a particular situation. For example, if a visually impaired user is speaking, the device can infer the degree of attention and distance of one or more people and deliver information only as needed. The user would want to know whether the other person shows interest in the conversation, which might be a cue to change or stop the conversation. If the other person leaves inadvertently, the device can warn the user that he/she was left alone and to stop talking. A similar situation can arise when a visually impaired user is waiting for a bus and asks another person to tell him when the bus arrives. If the person leaves without letting the user know, the system can detect that departure and inform the user.

Given the importance and incipient progress in social interaction applications, IPN researchers have begun to focus on assistive technologies that can detect head gestures, gaze direction, and the distance to those surrounding a visually impaired individual in a social situation. The ultimate goal is to discern nonverbal cues that reflect the degree of attention from other people so that the user can react in a timely and appropriate manner.

Although computer vision has potential as a basis for assistive technology, systems have yet to harness the full power of hardware advances and sophisticated algorithms. The current trend is to blend technologies to create devices and applications, as in Fifth Sense and Georgie, which exploit a range of smartphone technologies.

In this sense, computer vision–based assistive systems are at an intermediate stage. Growing past that stage will require advanced algorithms that can reliably interpret visual information, refining it enough to understand the content of an image or scene or a listener's attitude. Researchers must also create a general-purpose vision-to-touch translator or vision-to-audition translator that is sufficiently robust and dependable for everyday activities.

Above all, system developers must remember that assistive technology is user centered. Deep knowledge of a visually impaired user's needs and skills must guide the development from prototype conception to system testing to ongoing refinement. Only then can system functions meet user expectations. Each person has a different way of learning, feeling, and perceiving. Assistive technology devices must learn from their users and adapt parameters, particularly user feedback, to personalize the experience.

Many assistive techniques are appearing as smartphone applications. From the design perspective, this purely software solution greatly simplifies development, with the free-lunch promise of even faster processing in future devices. From the user perspective, the smartphone is ubiquitous and has no assistive device stigma. However, for some applications, the smartphone is not the best solution. For example, most users prefer the white cane to any ETA so far, not only because the cane costs little, is reliable, and does not need batteries, but also because users do not have to continually hold it where they want to detect obstacles—an often long and tedious process.

The basic properties of the white cane challenge designers to develop new ideas based on low-cost technology, state-of-the-art algorithms, and renewable energy sources.

Computer vision has indeed evolved swiftly, but it is still far from approaching the capabilities of human vision in interpreting scene content semantically. A combination of multiple technologies, such as computer vision, GPS, wireless Internet, and voice recognition, in a wearable platform similar to Google Glass can deliver a single, versatile, hands-free assistive device that can learn from the user and provide multiple functions. The feedback interface should be based on haptic and audio that do not interfere with hearing. The basic properties of the white cane challenge designers to develop new ideas based on low-cost technology, state-of-the-art algorithms, and renewable energy sources. Such ideas combined with miniaturized processors will surely hasten the time when computer vision has some of the nuances of human vision. C

Acknowledgments

This research was partially supported by SIP-IPN under grant 20121642. We thank Paul Riley of the US Peace Corps and the technical editors of *Computer* for their comments to improve our manuscript.

References

1. S. Schörnich et al., "Psychophysics of Human Echolocation," *Basic Aspects of Hearing: Physiology and Perception*, B.C.J. Moore et al., eds., Springer, 2013, pp. 3011–3019.

2. J.M. Loomis, R.L. Klatzky, and N.A. Giudice, "Sensory Substitution of Vision: Importance of Perceptual and Cognitive Processing," *Assistive Technology for Blindness and Low Vision*, R. Manduchi and S. Kurniawan, eds., CRC Press, 2013, pp. 161–192.
3. P. B. Meijer, "An Experimental System for Auditory Image Representations," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 2, 1992, pp. 112–121.
4. D. Dakopoulos and N.G. Bourbakis, "Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 40, no. 1, 2010, pp. 25–35.
5. A. Khan et al., "KinDetect: Kinect Detecting Objects," *Computers Helping People with Special Needs*, K. Miesenberger et al., eds., LNCS 7383, Springer, 2012, pp. 588–595.
6. P. Vera, D. Zenteno, and J. Salas, "A Smartphone-Based Virtual White Cane," *Pattern Analysis and Applications*, vol. 16, no. 1, 2013, pp. 1–10.
7. 7. V. Ivanchenko, J. Coughlan, and H. Shen, "Crosswatch: A Camera Phone System for Orienting Visually Impaired Pedestrians at Traffic Intersections," *Computers Helping People with Special Needs*, K. Miesenberger, et al. eds., LNCS 5105, Springer, 2008, pp. 1122–1128.
8. X. Yang and Y. Tian, "Robust Door Detection in Unfamiliar Environments by Combining Edge and Corner Features," *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (CVPR 10), 2010, pp. 57–64.
9. B.S. Tjan et al., "Digital Sign System for Indoor Wayfinding for the Visually Impaired," *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (CVPR 05), 2005, pp. 20–25.
10. J. Coughlan and R. Manduchi, "Functional Assessment of a Camera Phone-Based Wayfinding System Operated by Blind and Visually Impaired Users," *Int'l J. Artificial Intelligence Tools, Architectures, Languages, and Algorithms*, vol. 18, no. 3, 2009, pp. 379–397.
11. P.E. Lanigan et al. "Trinetra: Assistive Technologies for Grocery Shopping for the Blind." *Proc. Int'l Symp. Wearable Computers* (ISWC 06), 2006, pp. 147–148.
12. E. Tekin and J. M. Coughlan, "An Algorithm Enabling Blind Users to Find and Read Barcodes," *Proc. IEEE Workshop Applications of Computer Vision* (WACV 09), 2009, pp. 1–8.
13. T. Winlock, E. Christiansen, and S. Belongie, "Toward Real-Time Grocery Detection for the Visually Impaired," *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (CVPR 10), 2010, pp. 49–56.
14. C. Guida, D. Comanducci, and C. Colombo, "Automatic Bus Line Number Localization and Recognition on Mobile Phones—A Computer Vision Aid for the Visually Impaired," *Image Analysis and Processing*, G. Maino and G.L. Foresti, eds., LNCS 6979, 2011, pp. 323–332.
15. S. Krishna and S. Panchanathan, "Assistive Technologies as Effective Mediators in Interpersonal Social Interactions for Persons with Visual Disability," *Computers Helping People with Special Needs*, K. Miesenberger et al., eds., LNCS 6180, 2010, pp. 316–323.

Juan R. Terven *is a doctoral student at Instituto Politécnico Nacional, Mexico. He has been a visiting graduate student at MIT, a consultant on industrial embedded design, and a half-time lecturer at Instituto Tecnológico de Mazatlan, Mexico. Terven's research interests include embedded systems, computer vision, and assistive technologies design. Contact him at jrterven@mit.edu.*

Joaquín Salas *is a professor in the area of Computer Vision at Instituto Politécnico Nacional. His research interests include the development of assistive technology for the visually impaired and visual interpretation of human activity. Salas received a PhD in computer science from Monterrey Institute of Technology and Higher Studies (ITESM), México. He is member of IEEE. Contact him at jsalasr@ipn.mx.*

Bogdan Raducanu *is a senior researcher and project director at the Computer Vision Center in Barcelona, Spain. His research interests include computer vision, pattern recognition, and social computing. Raducanu received a PhD in computer science from the University of the Basque Country, Bilbao, Spain. Contact him at bogdan@cvc.uab.es.*

Selected CS articles and columns are available for free at http://ComputingNow.computer.org.