Camera-Based Access to Visual Information

J. Coughlan, Ph.D. and R. Manduchi, Ph.D.

Acknowledgments

Coughlan acknowledges support by the National Institutes of Health from grants No. 1 R21 EY021643-01, 2 R01EY018345-04, 1 R01 EY018210-01A1 and 1 R01 EY018890-01, and by the Department of Education, NIDRR grant number H133E110004.

Manduchi acknowledges support by the National Science Foundation (NSF) from grants No. IIS-0835645 and CNS-0709472, and by the National Institutes of Health under grant No. 1R21EY021643-01.

Introduction

Computer vision (also known as machine vision) is a form of artificial intelligence that strives to make computers able to "see" like normally sighted persons. The basic approach consists of analyzing visual information taken of a scene – either a static image or a video sequence, acquired by one or more cameras – and using software algorithms to infer important visual elements in the scene, including the presence, location and appearance of various objects and the 3-D scene layout. While computer vision is far from a solved problem, and is in fact the subject of intensive research, the last several decades of progress have led to a variety of successful algorithms, including OCR (optical character recognition), object recognition (including face

recognition), 3-D reconstruction of objects and other structures, and video analysis and event recognition. As we describe in more detail below, such algorithms drive a wide range of practical applications, many of them used in commercial products and equipment. Moreover, as computers become ever more powerful and compact, they make possible mobile platforms that deliver the power of computer vision to the fingertips of any smartphone (such as an iPhone or Android) owner.

The potential to harness computer vision to provide access to visual information that is otherwise inaccessible to visually impaired persons is extremely exciting and promising. However, important limitations of computer vision technology make it challenging to provide this kind of access. First, computer vision is successful in restricted domains but is often unreliable outside these domains, under the very types of highly variable and uncontrolled conditions that are commonly encountered by users "in the field." (For instance, OCR often fails unless the text to be read is a standard font that has been imaged clearly, which is often a problem for reading text printed on signs, which we discuss below.) Second, it is not yet capable of the truly high level of inference that would be desirable to many users – such as taking a picture of a scene and asking the computer "How do I get to the conference room?" or "Is there a place to sit nearby?" Finally, various hardware and processing constraints such as the limited field of view of the camera (which makes it difficult for a visually impaired user to locate an object of interest) and delays incurred by computer vision processing pose additional challenges to the design of an effective user interface for any computer vision-based assistive technology.

The aim of this chapter is thus to provide an overview of what computer vision does well, to

survey the applications that have been most successful for visually impaired users and to discuss the most important usability issues that must be considered in developing these applications. Similar material, but aimed more specifically at a computer science audience, is discussed in Manduchi and Coughlan (2011), and is the topic of an ongoing workshop series organized by the authors, Computer Vision Applications for Visual Impairment (CVAVI), which was previously held in 2005 in San Diego[1], 2008 in Marseille[2] and 2010 in San Francisco[3].

A Success Story: Optical Character Recognition

Some History

Without a doubt, the most successful image-based assistive technology to date is optical character recognition (OCR). An OCR system uses an imaging sensor to access text (typically, printed on paper) and some "intelligence" to translate the image content into letters and words. The output of OCR can then be recorded as a text file in various formats (e.g., PDF), and/or read aloud by a text-to-speech system. Thus, OCR allows a person with visual impairment to access printed information – a tremendous achievement, considering the paramount role of printed matter in human communication. In fact, OCR may also help persons who can see but have other forms of reading disabilities.

It is remarkable that the first fully electronic OCR system ever demonstrated was intended for use by blind people. The system was built in 1946 by a research group under the direction of

[1] http://users.soe.ucsc.edu/~manduchi/CVAVI/

[2] http://www.ski.org/Rehab/Coughlan_lab/General/CVAVI08.html

[3] http://italia.cse.ucsc.edu/~CVAVI10/

Flory and Pike at RCA Laboratories, with sponsorship by the Veteran Administration and the wartime Office of Scientific Research (Mann 1949). The "reading machine" would read text, spelling it out letter by letter; the machine was also able to read a few whole words. Its "eye" was a scanner moved over the text by hand, with 8 spots of light aligned in a column flashing on and off 600 times a second, and a photoreceptor that would read the light reflected off the printed paper. When illuminating a text character, only a few of these spots of light would be reflected by the paper; analysis of the pattern of reflected light by a system composed by more than 160 vacuum tubes produced the character, which would then be read aloud by activating one of 40 magnetic tape phonographs.

Although this initial prototype did not find practical use (due, among other things, to its cost and size), it spurred a whole new industry with important applications. The first OCR company, Intelligent Machines, was founded by Shepard and Cook in the early 50's. The Reading Machine, developed by Rabinow while at the National Bureau of Standards, showed improved performance through a "best match" procedure – essentially, by optically correlating each character of text against all possible alphanumerical characters. Image correlation (by digital means) has been a mainstream procedure for OCR for many years, especially for standard fonts. More recently, approaches based on pattern matching, for example via neural networks (Le Cun et al. 1990), have been used successfully in more complex situations, such as handwritten text. OCR quickly found extensive application in automatic document processing for business transactions and postal service (Hull et al. 1984). The development of accessible OCR systems with text-to-speech capabilities made this technology available to the visually impaired community. Two commercial products had a pivotal role in making OCR a widely used assistive

technology device: the Kurzweil Reading Machine and Arkenston's OpenBook.

In 1974, Kurzweil Computer Products developed the first commercial OCR that could recognize multiple fonts. (Previous systems used standardized fonts, such as the fixed-width mono-spaced format called OCR-A.) In his book "The Age of Spiritual Machines" (Kurzweil 2000), Ray Kurzweil recalls that an encounter with a blind person on a plane flight convinced him that the best use of this technology would be to support the blind community. The Kurzweil Reading Machine, which integrated omni-font OCR with a flat-bed scanner and a text-to-speech synthesizer, was introduced with much fanfare in 1976, and made commercially available in 1979. Extensive pre-production user testing supported by the National Federation of the Blind (NFB) were undertaken to ensure that this product would really be usable by visually impaired persons. Kurzweil Computer Products was later integrated in ScanSoft (a Xerox spin-off), which eventually merged with Nuance Communication.

Arkenstone, a non-profit organization founded in 1989 by Jim Fruchterman, also developed reading tools for people with disabilities. Arkenstone's reading system was originally based on technology by Calera Recognition System (an OCR company started by Fruchterman in 1982). It delivered OCR systems to more than 35,000 visually impaired individuals before it sold its business operations to Freedom Scientific in 2000. Arkenstone's OpenBook Scanning and Reading software (which utilizes both Nuance OmniPage and ABBYY FineReader OCR engines) is currently marketed by Freedom Scientific.

OCR by Sensory Substitution: The Optacon

OCR systems translate printed text into a computer-accessible format that can then be accessed via text-to-speech. One commercial product sidestepped the middle link of this chain, seeking to provide a blind user with more direct access to text via sensory substitution. The Optacon, developed by John Linvill and marketed by Telesensory Systems from 1971 to 1996, used a 24 by 6 pixel optical sensor that was manually moved across the text line to be read. A matching array of vibrating pins was used to "feel" the text, letter by letter, with one's fingertip, from the images taken by the optical sensor. (For more information about sensory substitution systems, see Chapter 8.)

The Optacon found success with a community of devoted users, who were not intimidated by the relatively long (two weeks) recommended training period. Typical reading speeds varied between 5 and 15 words per minute (Schoof 1975) – much lower than the typical Braille reading speed of approximately 125 words per minute. However, the Optacon provided an unprecedented level of access to printed text and to the nuances of typographic styles (Stein 1998).

Mobile OCR

While early OCR systems were very bulky and included a flatbed scanner to image the desired text and a desktop computer to perform the necessary software analysis, vast increases in computing power have recently enabled the development of portable OCR systems. As a result, a variety of OCR smartphone apps are now available for normally-sighted users, including the

ABBYY TextGrabber + Translator[4] and the Prizmo[5], as well as Word Lens[6], an "augmented reality" OCR app that reads all text visible in the camera's field of view, translates it to another language and graphically re-renders it on the viewfinder in place of the original text in the scene. The first commercial mobile OCR system designed for visually impaired users, the knfb Reader Classic, was released in 2005 by K–NFB Reading Technology, Inc.[7] (a joint venture between Kurzweil Technologies and the NFB). This handheld system consisted of a PDA (personal digital assistant) bundled with a separate digital camera, and was soon succeeded by smartphone versions, the knfbReader Mobile and kReader Mobile. Similar functionality is provided by the Intel Reader[8], which is a portable tablet device intended for a variety of special needs populations including the visually impaired and persons with dyslexia.

It is important to understand the great challenges that mobile OCR applications for visually impaired persons impose compared with standard desktop OCR, which was designed for use with high-quality images of printed documents obtained using a flatbed scanner. First and foremost, it may be difficult or impossible for the user to know where to point the camera so as to properly frame the text of interest in the camera's field of view. Indeed, the kReader Mobile/ knfbReader Mobile User Manual has an entire section on "Learning to Aim Your Reader," which includes instructions on practicing aiming the smartphone camera with a special training page. The aiming problem is especially severe if the user is unsure as to whether a sign (or other printed matter) is even present in the vicinity, as when searching a corridor for a particular room

[4] http://www.abbyy.com/textgrabber/
[5] http://www.creaceed.com/prizmo/iphone/
[6] http://questvisual.com/
[7] http://www.knfbreader.com/
[8] http://www.careinnovations.com/Products/Reader/Default.aspx

number.

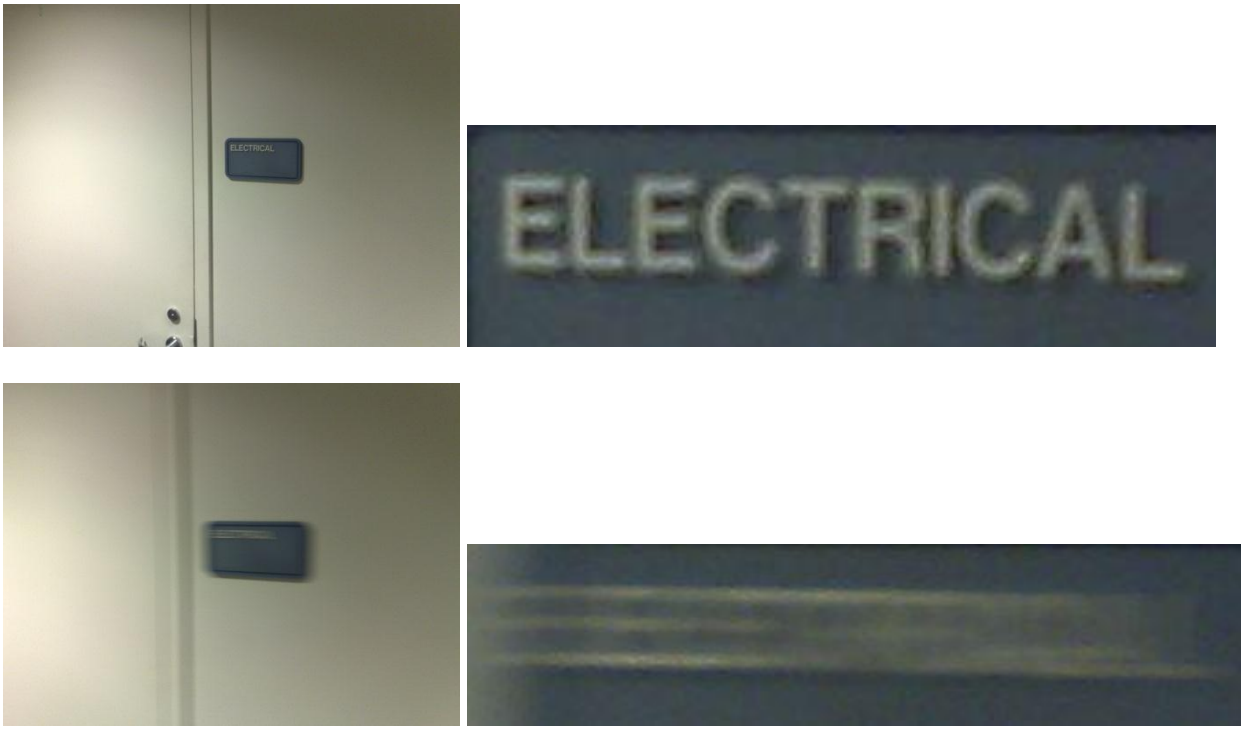

Figure 10.1. OCR and motion blur. Top row: clear smartphone camera image taken of indoor sign on left; zoomed-in portion on right shows that text is legible, and OCR reads it correctly. Bottom row: motion-blurred image taken by same camera on left; zoomed-in portion on right shows that text is unreadable, and OCR is unable to read it.

Second, even if the camera is pointed towards the desired text, if it is too close then the text may be partly cropped and/or out of focus; if it is too far then the text may be too small to read clearly, and/or the OCR may be confused by the expanse of non-text scene clutter surrounding the text. Third, images of text acquired by mobile devices are often poorly resolved because the text of interest is too far away, poorly illuminated, motion blurred (often exacerbated by a long camera exposure due to low light levels, see Figure 10.1) or appears on a curved surface (e.g.,

the label on a can of food). Finally, much text printed on informational signs is rendered in non-standard fonts – often combined with special symbols and commercial logos – which is difficult for OCR to read.

A large body of ongoing research is focused on addressing these challenges, illustrated in the following examples. The problem of detecting text in cluttered scenes is an increasingly popular topic of research in computer vision and document analysis (for samples of recent work, see Epshtein et al. 2010; Lee et al. 2011; Chen et al. 2011), including work on finding text that may be too small or poorly resolved to read (Sanketi et al. 2011). Related work (Coates et al. 2011; Wang et al. 2011) seeks approaches that integrate text detection with reading to improve performance on natural scenes. Finally, specialized techniques for detecting and reading text shown in LED and LCD displays – which are becoming increasingly common in household appliances such microwave ovens and DVD players – are being developed (Tekin et al. 2011) to provide visually impaired persons with access to these displays. Despite the considerable progress that has been accomplished in these areas, much work remains in making mobile OCR systems practical for use by visually impaired persons.

Other Semantic Information

The previous section on OCR focused on the recognition of printed text, which is the ubiquitous building block of an enormous range of documents and informational signs. However, many forms of semantic information are represented by visual patterns other than text, including standalone icons or signs primarily identified by the icons they bear (such as traffic signs and commercial signs), paper currency and barcodes, which cannot be read by OCR and are therefore

inaccessible to blind and visually impaired persons.

Reading Signs and Signals

Compared with OCR and text detection, relatively little research has focused specifically on the detection and recognition of non-text semantic information. Many important commercial signs (e.g., labeling stores and restaurants) are identified by distinctive non-text icons or logos, sometimes with accompanying text (which is often in a highly non-standard font that is difficult or impossible for OCR to read). Rather than attempting to recognize such signs individually and out of context, most work on sign recognition (Mattar et al., 2005; Silapachote et al., 2005) simplifies the problem by matching an image of an unknown sign to a database of known signs.

An important component of research on reading non-text signs addresses the need for robots and autonomous vehicles to recognize traffic signs and signals in order to safely navigate, negotiate roads and avoid contacting pedestrians. Of greater relevance to blind and visually impaired persons is the closely related problem of designing computer vision algorithms to find and recognize pedestrian signs and signals, including painted crosswalk patterns (Se 2000), Walk lights (Aranda et al. 2004) or other traffic lights (Park and Jeong 2009), which are of paramount importance for pedestrian safety at traffic intersections. As discussed in Chapter 3, crossing a street is a challenging (and dangerous) undertaking without sight. In particular, one needs to be well aware of the crosswalk layout and of the flow of traffic; figure out exactly where the crosswalk begins, and align himself or herself towards the correct crossing direction; estimate the precise time for crossing; and, once starting to walk, maintain the correct direction without drifting out of the crosswalk. Walk light timing information from Accessible Pedestrian Signals

is helpful in those intersections where they are installed (and some people can also use these signals to help align themselves correctly to the crosswalk), but unfortunately these signals are lacking at most intersections.

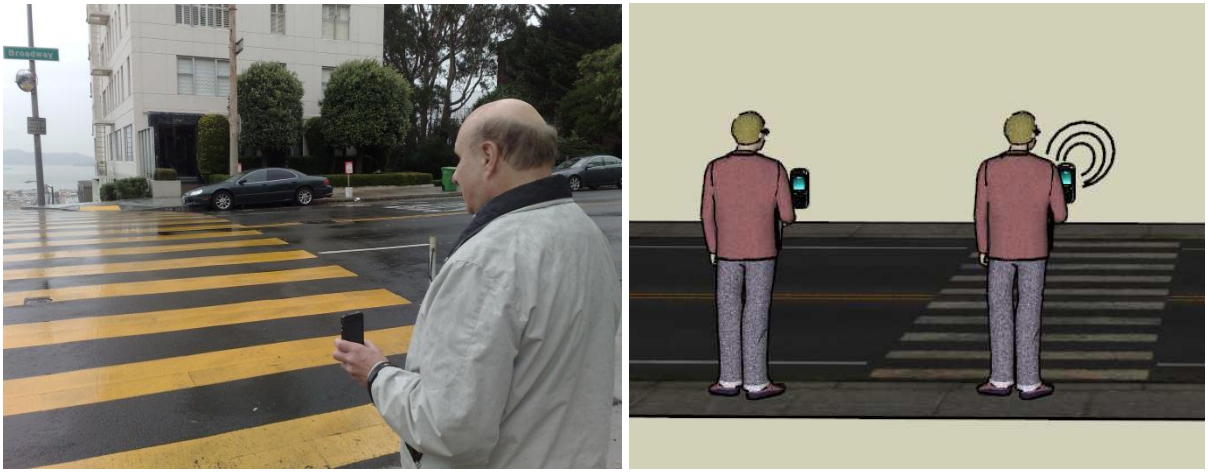

Figure 10.2. Crosswatch, computer vision-based smartphone system for providing guidance to visually impaired pedestrians at traffic intersections. Left: system in use by blind tester. Right: schematic shows how system provides audio cues to signal proper alignment of user to crosswalk.

Image-based technologies may be used to both estimate the precise geometry of the crosswalk (allowing one to align correctly with it), as well as to access information about the timing of the Walk lights, by extracting information from the signs, painted patterns and signals that are already present in traffic intersections. Prototypes of this kind of software have been ported to smartphones, including the "Crosswatch" system for recognizing crosswalks and Walk lights in real time (Ivanchenko et al., 2009; Ivanchenko et al., 2010), and tests with blind users have demonstrated the feasibility of this system. However, the challenge of performing the necessary pattern recognition reliably and swiftly, under a wide range of operating conditions (and

accommodating many variations in the appearance of pedestrian signs, signals and crosswalks), makes the solutions proposed so far unsuitable for wide adoption.

Barcodes

The last category of non-text semantic information we consider is the barcode. The most common barcodes are one-dimensional (1-D) patterns (such as the UPC, or "Universal Product Code," in widespread use in North America) labeling the vast majority of commercial products, which were designed to permit rapid product identification by a laser scanner. Several dedicated barcode readers have been designed for visually impaired users, such as the i.d. mate OMNI and Summit talking barcode readers (from Envision America)[9], but the growing power of smartphone technology has led to the development of barcode reader smartphone apps, such as the Red Laser iPhone app[10] intended for normally sighted users and the Digit-Eyes iPhone app[11] designed expressly for visually impaired users. As mentioned earlier, a major challenge imposed by any barcode reader is the difficulty of a blind or visually impaired user having to find the barcode on a package before it can be read; this difficulty motivated the development of smartphone barcode readers that interactively guide the user to the barcode (Tekin & Coughlan 2010; Kutiyanawala et al. 2011). An additional difficulty of the smartphone platform is the fact that the barcode is analyzed from an image taken by the smartphone camera, which is noisier and thus harder to decode than intensity data acquired using laser scanners (for which the 1-D barcode was originally designed); as a result, research (Gallo and Manduchi, 2011) has focused specifically on improving the accuracy of barcode recognition in challenging images.

---

[9] http://www.envisionamerica.com/products/idmate/

[10] http://redlaser.com/

[11] http://www.digit-eyes.com/

Two-dimensional (2-D) barcodes (such as the QR code and Data Matrix) were designed to be read by camera-based systems rather than laser scanners, and are superior to 1-D barcodes in that they are more easily read by camera-based systems and encode more data in a smaller area. Most modern smartphones come equipped with one or more apps capable of reading QR codes, which are increasingly used in magazines, tickets, signs, billboard advertisements and other printed matter to encode internet links to provide additional information about the printed content. While most QR codes are targeted at normally sighted smartphone users – who find it much less difficult to locate the codes and point their cameras accurately at them than do visually impaired users – there is ongoing work on using QR codes in applications intended for visually impaired persons, such as annotating objects in a museum and identifying product packages labeled with QR codes (Al-Khalifa 2008).

Signs for Wayfinding

Another intriguing application of non-textual signs is to support blind wayfinding. Sighted persons, when in unfamiliar places, routinely use available signage for understanding their location in the environment, and for determining how to reach a destination. Signs are particularly important in environments with complex layouts such as airports, where one needs to quickly find their way in stressful and potentially confusing situations. Unfortunately, signage is an inherently visual feature, and cannot be accessed without sight. (Braille or raised print signs are indeed accessible, but one first needs to physically reach them, which makes them useless for wayfinding in many situations.)

It is conceivable, though, that a camera may work as the "eye" of a blind person to detect existing signs. In fact, special signage that can be read efficiently by a machine may also be used for this purpose. For example, as mentioned earlier, two-dimensional bar codes (e.g. QR codes) are increasingly being used for smartphone-accessible information embedding. These signs are designed to pack as many bits of information as possible in a small amount of space. Similar types of signs have been used for wayfinding applications (Tjan et al. 2005). Another type of non-textual marker, with information embedded through colors, was developed by the authors (Coughlan and Manduchi 2009) and used for experiments in guided mobility (Manduchi et al. 2010; Figure 10.3). Color markers have the advantage that they require minimal computation for detection, and can be seen from a long distance even under relatively difficult light conditions (Bagherinia and Manduchi 2011).

Figure 10.3. Wayfinding experiments using a color marker that is easily detectable by a camera-equipped cell phone.

In some cases, camera-based sign detection can also be used to estimate one's location and

orientation ("pose") with respect to the sign. ARToolkit (Poupyrev et al. 2000), a 2-D barcode design standard with accompanying software developed at the University of Washington, has been employed extensively in "augmented reality" applications that use the estimated camera location to create special graphics effects visible through the viewfinder. The ability to compute and track one's relative position is an intriguing feature of image-based sign detection. This information could potentially be used to provide some local guidance to a blind person to reach a specific destination.

Other approaches currently being investigated for assisted wayfinding include embedding RFID tags in the environment (for example, under carpet tiles). RFID tags, which can be accessed by a suitable device at an appropriate distance, are thus equivalent to "signs" that can be placed throughout the environment (Kulyukin et al. 2004; Ross et al. 2010).

Camera-accessible signage is a promising approach to provide location-based information to persons without sight. It represents a sensible and cost-effective way to render information perceptible to all – regardless of whether one can see or not. As such, it fits well within the framework of "Universal Design", a concept that has gained increasing popularity over the past decade. Universal Design (Preiser and Ostroff 2001) aims to produce buildings, products and environments that can be used by all (with or without disabilities), rather than having to be adapted to one's particular needs. One of the principles of Universal Design is "Perceptible Information," which states that the design should communicate "necessary information effectively to the user, regardless of ambient conditions or the user's sensory abilities." Signage that can be accessed via mobile imaging represents one sensible way to achieve this goal.

Computer Vision for Mobility

As discussed at length in Chapter 3, mobility (the ability to move around safely and efficiently) and orientation (the ability to find a path to destination) are critical skills for a visually impaired traveler. The long cane and the dog guide are the standard mobility tools. Several ultrasonic or light-based mobility devices (Electronic Travel Aids or ETAs) have been proposed in the past; their features and limitations are described in Chapter 3. It is only natural that researchers would consider applying image-based techniques to orientation and mobility.

Much of this effort derived from experience acquired research in autonomous robotics. After all, an autonomous robot is a "blind" agent: it needs sensors for obstacle detection, along with some level of intelligence for path planning and navigation. Reliable autonomous navigation with both wheeled and legged platforms has been demonstrated in recent years, owing much to the use of imaging and depth sensors. It would be a mistake, though, to think that by simply augmenting a sensory system with some sort of acoustic or tactile interface, a usable mobility tool for a blind traveler would be created. Humans are not robots; what works well on an autonomous navigation platform may be totally inadequate for a blind traveler. The reader is referred to the "Recommendations for the design process" section in Chapter 8 for a thorough discussion of the "engineering trap" – developing technical solutions without proper consideration of the end user, which all too often results in lack of adoption.

Vision-Based Electronic Travel Aids

A number of vision-based ETA prototypes have been developed in recent years. It should be noted that very few of these systems have undergone thorough testing with visually impaired users in realistic situations. Hence, the following exposition should be considered more as an indication of recent research trends, rather than a list of tools ready for adoption.

Sensory systems for mobility support are normally designed to detect and possibly characterize environmental features that are important for safe ambulation: different types of obstacles, steps or curbs, or points of access (doors or passages). In most cases, detection is based on depth perception, achieved either by stereo vision or by active triangulation. Stereo vision is based on the triangulation principle: if a surface point is seen by two cameras at a certain distance from each other, it projects onto two different locations in the two cameras' focal planes. The distance between the locations of the projections ("disparity") is inversely proportional to the distance of the point from the cameras. Thus, by matching each point in the image produced by one camera to the corresponding point in the image produced by the other camera, one may obtain a "depth image" – effectively, a three-dimensional representation of the whole scene (see Figure 10.4).

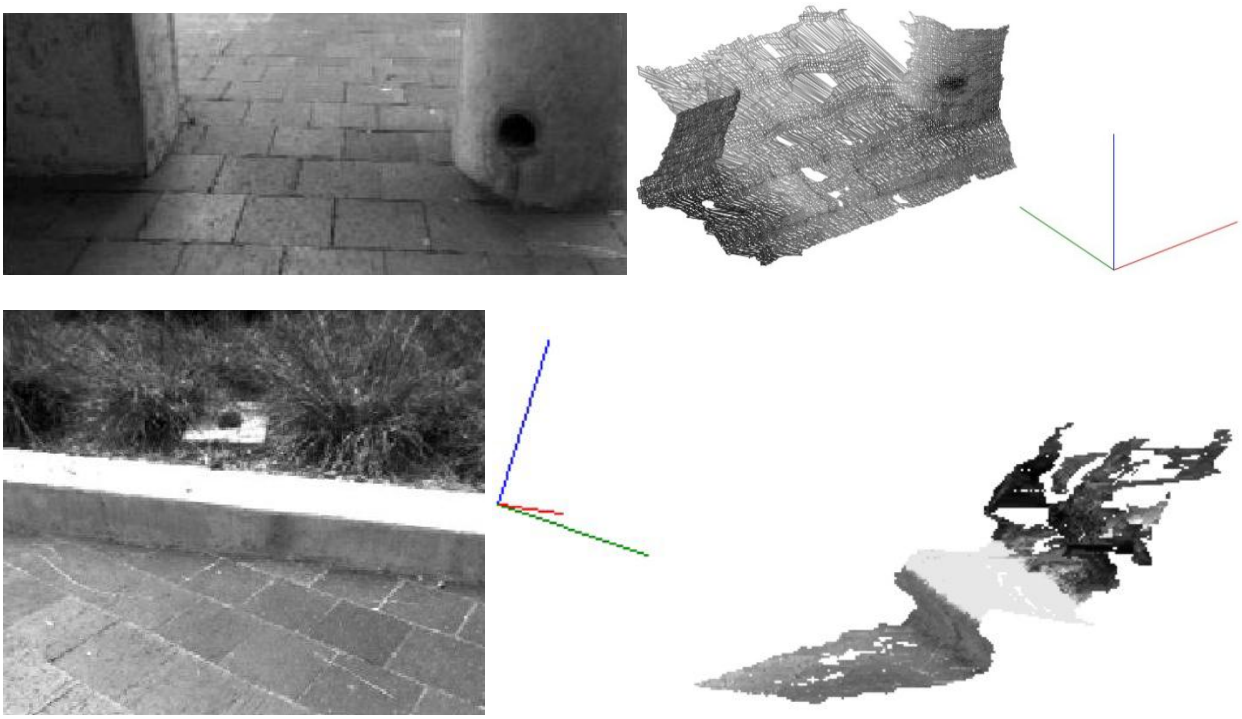

Figure 10.4. Stereo camera system estimates 3-D structure of scenes. Top row: scene shown on left, 3-D reconstruction on right clearly shows wall and column protruding from ground plane (xyz-axes drawn with green line indicating the camera's line of sight). Bottom row: scene of curb shown on left, 3-D reconstruction shows step structure, which is clearly present despite distortion of reconstruction.

Active triangulation systems operate on a similar principle, but do not need two cameras. Instead, light is produced by a source (typically, a laser or LED source) that is geometrically calibrated with the camera. By locating the return of the light reflected by a surface point on the image, the depth of this point is estimated. Active triangulation systems may use a single point source (producing only one depth measurement), a "striper" light source (producing readings over a plane in space), or a pattern of points that sample the space (as with the Microsoft

Kinect[12] system). Although stereo and active triangulation systems produce the same type of data, there are important technical differences between the two. Stereo requires the visible surfaces to reflect light from a light source (and thus cannot work in the dark), and doesn't work well with untextured surfaces (e.g., a white wall), because it is difficult to match points in the two images in these situations. Active triangulation, conversely, does not work well under full sunlight, because the light from the environment overwhelms the light irradiated by the source, thus making detection of the return from the light source difficult. Both systems fail in front of a transparent surface such as a glass door.

Other types of depth sensors are based on computing the "time of flight" (TOF), i.e., the time that it takes for a pulse of light (or a modulated light signal) emitted by a source to reach a surface in the scene, reflect back from the surface, and return to the sensor. The distance to the surface is thus proportional to the time of flight. Systems that use the time of flight principle (LIDARS) require that the light source be mechanically rotated to span the desired section of the scene. Imaging TOF systems have recently been marketed; these "3-D cameras" have great potential for mobility applications.

An example of an ETA based on active triangulation is the "Teletact" system (Farcy et al, 2006). This device, normally clipped on a long cane, measures the distance to a surface using a single point light source (a laser pointer), and communicates it to the user using a simple tactile or acoustic interface. The device can reliably measure distances up to 6 meters with a very narrow "field of view" angle. According to Farcy at al. (2006), Teletact has been tested with more than

[12] http://www.xbox.com/en-US/kinect

200 users. Note that the narrow field of view angle achieved by the laser beam represents an important difference with respect to the ultrasonic ETAs mentioned in Chapter 3, which typically have a much broader receptive field.

In general, safe ambulation requires awareness of multiple environment features, such as steps or drop-offs that, if undetected, could lead to a fall. Recognizing these features by means of point distance measurements may be difficult or impossible, and thus other techniques that use multiple measurements or depth imaging may be needed. For example, Yuan and Manduchi (2004) developed a prototype active triangulation system that not only measures distance, but also analyzes the time profile of distance as the user moves the device (pivoting it around a horizontal axis) to detect depth discontinuities that could signify steps or drop-offs. This system was later extended to use a laser striper, which allows for detection of surface features from a single image (Ilstrup and Manduchi 2011).

Stereo-based detection of curbs and steps for blind navigation was first proposed by Molton et al. (1998). If three-dimensional information about the environment is available (from a stereo camera or other depth imaging device such as the Kinect), this data can be integrated through time using a technique called "simultaneous localization and mapping" (SLAM). This allows for the geometric reconstruction of the environment, and at the same time enables self-localization. This technique was recently proposed as a means to support blind mobility (Pradeep et al. 2010). Image-based techniques can produce a very rich representation of the environment geometry, which makes it possible to identify features of interest for safe mobility. This approach may also support orientation, by computing paths to destination. One should be very careful, however,

when considering how this technology can be used to effectively help a visually impaired traveler. As discussed at length in Chapter 8, communicating geometric descriptions to a person without sight can be very challenging, and there is no clear agreement as yet about what level of information should be communicated (whether only high-level, semantic information, or some sort of general description), and what type of interface should be used (speech, sound, spatialized sound, or tactile).

But the main challenge for a developer of assistive technology for mobility is to prove that a proposed new device is superior to the baseline tool – the long cane. The cane is economical; it always works and never runs out of power; when not in use, it can be folded and put in one's pocket; and it signals the presence of a blind person to other people. Expert users of the long cane are naturally reluctant to the idea of giving it up for something that is less tried and tested (or even augmenting it with any kind of new technology). Thus, a developer may have to work hard (and provide strong experimental results with adequate user studies – see Chapters 6 and 7) to convince more than a few enthusiastic "early adopters" to convert to a new technology.

It should be noted, however, that the long cane cannot detect all types of obstacles. In particular, the user is not protected against obstacles at head height. In a survey with 300 blind and legally blind individuals (Manduchi and Kurniawan 2011), 13% of the respondents reported experiencing collision with head-level obstacles at least once a month, often with traumatic consequences. Hence, technology that could reliably detect obstacles at head height (Jameson and Manduchi 2010) may have good potential for acceptance at least by a portion of blind and low vision travelers.

Recognizing Stuff

It is often difficult or impossible for a blind or visually impaired person to determine the identity of an object even when it is within reach; in some circumstances, it may be inconvenient to seek this kind of information by simply asking someone for help, as when a blind person wishes to learn who is standing nearby at a party or other gathering. For such problems, computer vision object recognition algorithms, which recognize an object in an image as an instance of an object category, are a promising tool.

Object recognition is a fundamental problem in computer vision (see Szeliski 2010 for a recent overview), and has become powerful enough for widespread use in visual online search engines. Many smartphone apps enabled by object recognition have recently emerged. Perhaps the best known example is Google Goggles[13], a smartphone app which automatically recognizes many types of objects, including landmarks, book covers and artwork as well as text and logos. Other similar apps include oMoby, A9's SnapTell and Microsoft's Bing Mobile, most of which are optimized for commercial product visual searches. These apps are intended for normally sighted users, who can easily center the object of interest in the camera viewfinder, snap a photo and wait for a response. Unfortunately, such a user interface is impractical for many blind and visually impaired persons, who may have great difficulty knowing which objects are visible to the camera – and in some cases may not even know whether an object of interest is even present. For this reason, few object recognition systems are used by this population (unless one categorizes OCR as a type of object recognition).

[13] http://www.google.com/mobile/goggles/

Some important exceptions to this general rule are object recognition systems that employ user interfaces tailored to the needs of blind and visually impaired users, typically focusing on a specific problem domain rather than attempting completely general object recognition. One such domain is the problem of grocery product identification, which Winlock et al. (2010) approach by matching the packaging of an unknown product – which can include any combination of text, logos and other graphics – against a database of known products. This approach circumvents the problem of attempting to identify a product solely through OCR, which is unreliable for the types of non-standard fonts (and inhomogeneous backgrounds) that are typically printed on grocery products; moreover, unlike product identification based on barcodes, it doesn't force the user to locate the barcode on a package, which can be a difficult task in itself. The user interface in Winlock et al.'s system also alleviates the problem of knowing where to aim the camera through the use of a "mosaicing" technique that combines multiple views of the shelves taken in video over time to assemble a single coherent image of the entire scene.

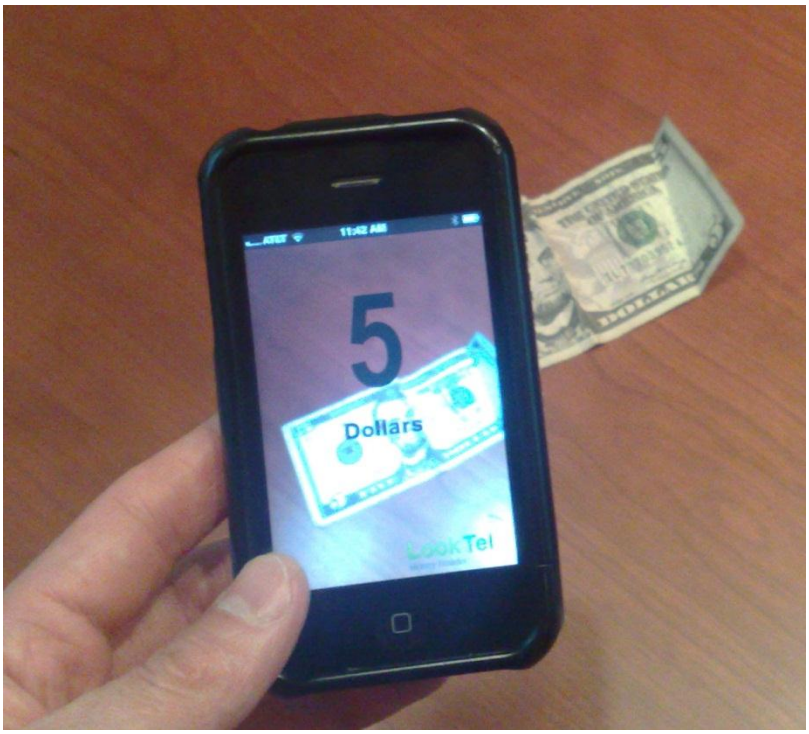

Figure 10.5. LookTel Money Reader, a currency reader app for the iPhone. System detects and reads aloud the denomination of the bill (in this case, $5) in real time.

Another object recognition-based system tailored to the needs of blind and visually impaired users is the currency reader, which determines the denomination of American paper currency (unlike paper bills in other countries, whose denominations may be distinguished simply by the size of the bill) – information that is vitally important to a visually impaired person anytime he/she conducts a cash transaction. Standalone currency readers have been available for many years (such as the Note Teller[14]), but smartphone-based currency readers (Liu, 2008) have become commercially available, including the knfb Reader Mobile and the LookTel Money Reader[15] smartphone app (see Figure 10.5). The LookTel system is distinguished by its ability to

[14] http://www.brytech.com/noteteller

[15] http://www.looktel.com

recognize currency in real time, as the user is aiming the camera towards the bill, rather than having the user take a photo, wait for the results to be read aloud, and possibly repeat the process until a suitable photo has been taken.

The ability to recognize faces is another specific object recognition task that would be very valuable for blind and visually impaired persons in some social contexts. Not only is it awkward to have to ask people around you who is near, but a partially sighted person whose visual impairment is not immediately obvious to others may risk offending people by failing to recognize them. In addition, people with otherwise normal vision who suffer from prosopagnosia (face blindness) could also benefit from a face recognition system. Face recognition is an active topic of research in computer vision (see Li and Jain 2005 for an overview) that is especially challenging because of the enormous variability that a person's facial appearance undergoes due to varying facial expressions, lighting conditions, viewing angles and the placement of hair, glasses, clothing, etc. Because of these challenges, relatively little research has been done on practical face recognition systems for blind and visually impaired persons. However, Kramer et al. (2010) implemented a successful prototype smartphone-based face recognition system that announces with text-to-speech the names of people identified by the system. Going beyond basic face recognition, other research (Krishna et al. 2005; Gade et al. 2009) explores a wearable interaction assistant system to identify and interpret facial expressions, emotions and gestures, which are another important form of information (in addition to identity) that may be inaccessible to people with vision loss.

Finally, in some circumstances a blind or visually impaired person (possibly someone with color

blindness) may know the identity of an object but not its color, as when choosing an article of clothing to wear or selecting a color-coded file folder. In such cases a system that automatically recognizes color can be helpful. Color identifiers (also known as color recognizers) such as the ColorTest (from Caretec), Brytech's Color Teller or Cobolt Systems' Speechmaster speak aloud the color of a surface that the device is pointed to; not surprisingly, a 'Color Identifier' app[16] is now available for the iPhone with similar functionality. Research on the related but more difficult problem of matching patterns according to both color and texture may be useful for helping visually impaired people match multiple items of clothes (Yang et al 2011).

Image Enhancement for Low Vision

In previous sections we have emphasized assistive technology systems that harness computer vision to provide audio or tactile feedback about a visually impaired person's environment. Some people with sufficient partial vision, however, may prefer user interfaces that make the most of their residual vision by using computer vision or image processing to *enhance* images of the scene, which are presented on a computer/smartphone screen or via a head-mounted display. As discussed in Chapters 2 and 4, there are many different types of functional vision loss, and we will outline approaches to image enhancement according to some of the more common types.

A relatively minor form of vision impairment is color blindness, which is a reduced ability to perceive color or color differences (and which often occurs in people with otherwise normal vision). As an alternative to the types of color identifier systems discussed at the end of the

[16] http://www.greengar.com/apps/color-identifier/

previous section, which announce (using text-to-speech) the color pointed to by a camera or similar light probe, the colors in an entire image (or electronic document) can be re-mapped to a new color palette and re-rendered, so as to make the colors appear as perceptually distinct as possible to the viewer while preserving the approximate contrast between different colors that would be perceived by a normally sighted viewer. (The most appropriate mapping depends not only on the range of colors present in the image but also the particular type of color blindness.) "Post-publication" techniques (Jefferson and Harvey 2006) address the problem of how to re-render a static image, webpage or entire document on a computer screen or printed page. The same approach can also be applied to video, and when it is implemented in real time it constitutes an augmented reality system, such as the DanKam[17], an experimental app iPhone app (designed for the most common form of color blindness, anomalous trichromancy) which continuously re-renders the scene acquired by the camera on the smartphone's viewfinder.

Much more serious functional vision impairments include poor acuity, poor contrast sensitivity and tunnel vision, which arise from multiple causes. Poor acuity is a particularly debilitating problem because it can impair a person's ability to read and to find and recognize objects, even under good lighting conditions and at high contrast. Magnification is a standard technique for addressing this problem (Wiener et al. 2010), either using an optical telescope (for distant targets) or a screen magnifier (for print). Unfortunately, as discussed in Chapter 4, a fundamental limitation of magnification is that it reduces the field of view – in effect, causing a type of functional tunnel vision, which makes it harder to find a target of interest, or to interpret an extended pattern larger than the field of view. (The problem is exacerbated with poorer acuity,

[17] http://dankaminsky.com/dankam/

which requires the use of higher magnification.) One computer vision-based attempt to mitigate this problem in the context of reading signs is the "Smart Telescope" SBIR project from Blindsight Corporation[18], which automatically detects text regions in a scene acquired by a wearable camera and presents the regions one at a time to a partially sighted user, using a head-mounted display that zooms into the text to enable him/her to read it.

Many systems that magnify scenes for people with poor acuity also enhance contrast (Harper et al. 1999), such as the head-mounted system created by Li et al (2011), since poor acuity due to central field loss is often associated with low contrast sensitivity. While simple contrast enhancement techniques are often adequate for improving the readability of text (typically characterized by a single foreground color against a single background color) – techniques that are available in many video magnifiers (Sardegna and Paul 1991) – enhancing the visibility of an entire natural scene with a multitude of objects and textures is a complex problem. Eli Peli and his collaborators have investigated this problem using a variety of contrast enhancement techniques, including narrowband contrast enhancement (Peli and Peli 1984), contrast enhancement in the JPEG and MPEG domains (Tang et al. 2004; Kim et al. 2004) in which images and video are compressed and wideband enhancement (Peli et al. 2004), which adds highly visible edge and bar feature contours to the image.

Finally, the restricted field of view associated with tunnel vision has inspired the development of novel techniques for processing and displaying images, most involving some form of minification (the opposite of magnification, used to compact a large field of view into the much

[18] http://www.blindsight.com/

narrower field of view of the person's eyes). A promising approach due to Peli (2001) uses a head-mounted display to superimpose minified edge images of the entire scene over the user's natural vision. The key to the approach is the fact that the edge pixels comprise a small fraction of pixels in the original images, so that the superimposed edge image interferes minimally with what the user already sees unaided. Experiments with tunnel vision patients showed that the system increased the speed with which they are able to conduct search tasks (for targets outside their natural field of view).

Many experiments have demonstrated statistically significant benefits to users of these types of image enhancement systems to perform real-world tasks (Luo and Peili 2011). However, for many applications such as image enhancement for viewing TV and video (Peli and Woods 2009), the determination of quantitative measures to objectively evaluate the benefits of image enhancement is a major challenge in itself. Indeed, given the extremely variable and multi-dimensional nature of functional vision deficits among persons with visual impairments, coupled with the highly individual needs of this population, this challenge is likely to be a research enterprise as great as the development and improvement of the image enhancement techniques themselves.

Usability

There are several aspects of a camera-based system that need to be addressed before it can be considered usable by visually impaired persons. In the following we discuss a few issues related to the physical placement of the system, the quality of the task being performed, the communication of relevant data to the user, and the distribution of processing tasks between the

local platform and remote (or even human) resources.

Wearable or Portable?

One important issue is how exactly the visually impaired user will carry the camera or cameras. The camera needs to have an adequate clear field of view, and thus cannot be placed anywhere the view could be occluded. At the same time, cosmetic considerations may discourage the placement of cameras in highly conspicuous locations. The availability of miniaturized cameras, which could be embedded, for example, in someone's eyeglasses, may help reduce some of these concerns. An alternative solution to "wearable" cameras is the use of a hand-held device such as a camera-equipped smartphone. Using a widely available commodity such as a smartphone has many advantages with respect to customized solutions, in terms of cost, availability of support, and convenience. In addition, a smartphone carries none of the stigma normally attached to assistive technology devices. Assistive technology "apps" are currently being created for both the Android and iPhone platforms. Of course, all aspects of accessibility need to be considered: if a blind user cannot start or control the application on the smartphone because this requires visual feedback, then even the most powerful application would be useless! Thankfully, accessibility interfaces (e.g. VoiceOver for Apple devices) are available to enable non-visual interaction even with touchscreen-based smartphones (see also Chapter 11).

In spite of their attractive features mentioned above, hand-held smartphones may not be the most desirable choice for all applications. A smartphone is ideal for things like checking the value of a banknote or for reading a bill or a menu via OCR and text-to-speech. However, holding a smartphone by hand and pointing it around for an extended period of time while walking in

search of a sign (indicating, for example, the location of a restroom or of the elevator) may be unwieldy, especially considering that the user would normally have one hand already occupied with handling the long cane or holding the dog guide. For this type of application, a wearable camera system that does not need to be constantly held may prove a more convenient solution.

Performance

Simply stated, if an assistive technology solution doesn't work well enough, it won't be used. How to quantify "acceptable performance," however, is not always clear.

A camera-based system can be characterized by a number of parameters, including: the field of view of the camera; the image resolution (the number of pixels in the image); the effective frame rate (the number of images that can be processed per second). In addition, there are application-specific performance measures. Consider for example a device used to find a particular sign in the environment. This system may be evaluated in terms of its detection rate (the percentage of times the sought sign is correctly detected in an image) and of its false positive rate (the percentage of times that the system declares a detection when in fact there was no sign visible).

Detection and false positive rates are normally related to each other via specific parameters in the algorithms (e.g., a threshold in the classification algorithm). By tweaking such parameters, one may be able to increase (or decrease) the detection rate, but this will necessarily come at the cost of an increased (or decreased) false positive rate. Fixing the "operating point" of the system (that is, tuning the parameters to obtain a certain detection / false positive rate pair) is a difficult art. One may, for example, decide on a maximum acceptable value for the false positive rate, under

the assumption that if the system produces too many false alarms, the user will decide to turn it off.

The resulting detection rate depends on a number of factors, including the size of the sign, the distance at which it is seen, the field of view of the camera, and the image resolution. A sign of a certain size, seen at a certain distance, will be easier to detect if the image resolution is high or if the field of view angle is small (e.g. using a telephoto lens). Increasing the image resolution, however, typically leads to longer processing time, which may reduce the effective frame rate. By contrast, a narrow field of view implies that a smaller portion of the environment is visualized at each frame. Low frame rate and narrow field of view may both hamper the visual exploration task if one is using the system to "scan" the environment in search of the sign, perhaps by rotating the camera around a vertical axis (see Figure 10.3). If the camera is moved too fast, there may be portions of the visible space that will not be processed because they happen to appear between two consecutive frames. Thus, the user may have to apply extra attention while exploring the space with the camera, possibly repeating the scanning operation several times to make sure that the entire scene is correctly analyzed.

It should also be noted that fast camera motion may produce blurred pictures and thus complicate processing by the computer vision algorithm. This effect is particularly noticeable with low ambient light, as in this case the system needs to increase the exposure time of each image, which is liable to increase the risk of motion-induced blur.

As this simple example suggests, the performance of a camera-based system is a function of

multiple interconnected components. Ultimately, the system should accomplish the task it was designed for in a reasonable amount of time and without requiring too much effort of the user. Optimizing the system for best performance, and quantifying the overall quality of the user-mediated application, is a difficult but critical component of the design process.

User Interface

A camera-based system implicitly performs some sort of sensory substitution. The visual data is "digested" by the computer vision algorithm, and the output of this processing (be it the detection of a step two meters away, the OCR decoding of a piece of text, the brand of a can of food, or the presence of a sign on a wall) is communicated to the user via one or more of his or her remaining senses. Clearly, the modality for this communication should be application-dependent. In some cases, a few bits of information is all that is needed (for example, to communicate the presence of a head-level hazard in time for the user to stop and avoid it). In other cases (e.g., OCR access to text), more data needs to be communicated, in this example via text-to-speech.

Specific to the case of camera-based systems are situations in which a stream of geometry-related information needs to be communicated to the user as he or she is maneuvering the camera. Consider for example the case of a smartphone used for mobile access to text, for example to read a bill received in the mail. As mentioned earlier, one of the challenges of using such a system is that the smartphone needs to be positioned so that the camera has a clear view of the entirety of the text (an operation that may be challenging without sight). The system may provide continuous information to the user as to how to move the camera to improve framing of

the desired text (e.g. up/down, left/right, closer/farther). As another example, a wayfinding application may inform the user of the distance and bearing angle of a detected sign.

There is no general rule about what combination of speech, sound, spatialized sound (if the user is wearing headphones) or tactile signal (vibration) is preferable for communicating geometry-related information, in part because each application has its own specific requirements. Precious little published research has performed comparative evaluation of multiple interfaces in this context (Ross and Blasch 2000; Marston et al. 2006; Walker and Lindsay 2005). Indeed, user interface design often winds up being "the afterthought at the end of the project" (Miele 2005), receiving less attention than other technical components of the device. Yet, the interface, from the user's standpoint, is one of the most important aspects of the sensory substitution system: a poorly designed interface may make an otherwise impeccable algorithm practically unusable.

From Local to the Cloud to the Crowd

As pointed out earlier, the processing speed (measured, for example, by the effective frame rate) is a critical component of a camera-based system. There is no universal rule for what the minimum acceptable response time of a system should be. For example, if a smartphone is used to determine the brand of a can of food, it may be acceptable to wait for a few seconds from the time a snapshot is taken to the time the response is uttered by the text-to-speech system. The situation is different if the computer vision system is supposed to provide a stream of signals, for example to help the user point the camera correctly in order to take a well-framed picture that will then be OCR-processed. In this case, a system that takes more than one or two seconds to process an image at each individual camera pose may be cumbersome to operate. Other

applications (e.g., detection of a particular sign or of a hazard as one is walking) may have even more stringent frame rate requirements.

Image processing is notoriously time-consuming. An image is composed of hundreds of thousands or millions of pixels, and the computer vision algorithm may require multiple iterations of complex routines involving possibly large neighborhoods of each pixel. Miniaturized platforms (e.g., smartphones), in spite of technological innovations such as multiple core processors and graphics processing units (GPU), are not as powerful as desktop computers. Thus, in order to increase processing speed, it may be sensible to use remote servers ("the cloud") to perform operations that would require too much time on the local processor. Of course, this requires that a good wireless Internet connection be available between the local platform (e.g., the smartphone) and the remote server. Indeed, transmitting streams of high-resolution images to the server may wind up taking more time than performing some simple local processing! Thus, a more effective strategy may consist of distributing the processing load between the local processor and the remote server. For example, input images may be analyzed locally to extract relevant features, which are then transmitted (using much less data than the original image) to the remote server for further analysis (Chandrasekhar et al. 2009).

Another intriguing possibility offered by the increasing availability of ubiquitous wireless connection is to do away with computer vision altogether, and rely instead on input from remote human assistants looking at images taken by the blind user and sent to them via Internet. One example of this approach is Sight On Call, a product being developed at Blindsight with funding from the NIH. Sight On Call is an on-demand assistance service for blind and moderate low-

vision persons. The user of this service contacts specially trained operators who, based on sensor data (e.g., GPS location) and images taken by the user's cell phone, can provide specific assistance as regards wayfinding and object recognition. A different approach is taken by VizWiz, developed by Bigham's group at the University of Rochester (Jayant and Bighan 2010). VizWiz uses "crowdsourcing" mechanisms (specifically, Amazon's Mechanical Turk[19]) to enable the user to ask and get answers to queries about an image taken with his or her cell phone.

Using "human intelligence" rather than computer vision is attractive for multiple reasons. Humans are much more reliable than computers for most tasks that require image analysis. More important is the fact that humans can answer complex and generic questions such as "Where am I?" or "What is close to me?" At the same time, this approach has some intrinsic limitations in terms of latency. The image taken by the user's cell phone needs to be transmitted to a server and analyzed by one (or more) human operators, before the answer to the user's query is sent back and uttered via text-to-speech. This chain of operations may take up to a few seconds. As mentioned earlier, this delay may be irrelevant for some applications, yet unacceptable for other tasks that require close to real-time feedback.

References

Al-Khalifa, H. 2008. Utilizing QR Code and Mobile Phones for Blinds and Visually Impaired People. In Proc. *International Conference on Computers Helping People with Special Needs (ICCHP '08)*. Linz, Austria.

[19] http://www.mturk.com

Aranda, J. and Mares, P. 2004. Visual System to Help Blind People to Cross the Street. In Proc. *International Conference on Computers Helping People with Special Needs (ICCHP '04)*. Paris, France.

Bagherinia, H. and Manduchi, R. 2011. Robust real-time detection of multi-color markers on a cell phone. *Journal of Real-Time Image Processing*, 6.

Chandrasekhar, V., Takacs, G., Chen, D. M., Tsai, S. S., Grzeszczuk, R., and Girod, B. 2009. CHoG: Compressed histogram of gradients - A low bit rate feature descriptor. In Proc. *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

Chen, H., Tsai, S. S., Schroth, G., Chen, D. M., Grzeszczuk, R. and Girod, B. 2011. Robust text detection in natural images with edge-enhanced maximally stable extremal regions. In *Proc. IEEE International Conference on Image Processing (ICIP 2011)*.

Coates, A., Carpenter, B., Case, C., Satheesh, S., Suresh, B., Wang, T. and Ng, A.Y. 2011.Text Detection and Character Recognition in Scene Images with Unsupervised Feature Learning. In Proc. *International Conference on Document Analysis and Recognition (ICDAR 2011)*.

Coughlan, J. and Manduchi, R. 2009. Functional assessment of a camera phone-based wayfinding system operated by blind and visually impaired users. *International Journal on Artificial Intelligence Tools, Special Issue on Artificial Intelligence Based Assistive*

Technologies: Methods and Systems for People with Disabilities, 18:379-397.

Epshtein, B., Ofek, E. and Wexler, Y. 2010. Detecting Text in Natural Scenes with Stroke Width Transform. In *Proc. Conference on Computer Vision and Pattern Recognition (CVPR 2010)*.

Gade, L., Krisha, S., Panchanathan, S. 2009. Person localization using a wearable camera towards enhancing social interactions for individuals with visual impairment. In *Proc. of the 1st ACM SIGMM international workshop on Media studies and implementations that help improving access to disabled users* (MSIADU '09). ACM, New York, NY, USA, 53-62.

Gallo, O. and Manduch, R. 2011. Reading 1-D Barcodes with Mobile Phones Using Deformable Templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*. 33(9).

Harper, R., Culham, L. and Dickinson, C. 1999. Head mounted video magnification devices for low vision rehabilitation: a comparison with existing technology. In *Br. J. Ophthalmol.* , 83(4), 495-500.

Hull et al. 1984. Hull, J.J, Krishnan, G., Palumbo, P. , and Srihari, S.N. Optical character recognition in mail sorting: A review of algorithms. Technical Report 214, CS Department, State University of New York.

Ilstrup, D., and Manduchi, R. 2011. Return detection for outdoor active triangulation. In *Proc. SPIE 3D Image Processing (3DIP) and Applications.*

Ivanchenko, V., Coughlan, J. and Shen, H. 2009. Staying in the Crosswalk: A System for Guiding Visually Impaired Pedestrians at Traffic Intersections. In Proc. *Association for the Advancement of Assistive Technology in Europe (AAATE 2009)*. Florence, Italy.

Ivanchenko, V., Coughlan, J. and Shen, H. 2010. Real-Time Walk Light Detection with a Mobile Phone. In Proc. *International Conference on Computers Helping People with Special Needs (ICCHP '10)*. Vienna, Austria.

Jameson, B., and Manduchi, R. 2010. Watch your head: A wearable collision warning sensor system for the blind. *IEEE Sensors 2010 Conference.*

Jayant, C. and Bigham, J. 2010. VizWiz::LocateIt — Enabling blind people to locate objects in their environment. In *Proc. IEEE Workshop on Computer Vision Applications for the Visually Impaired.*

Jefferson, L. and Harvey, R. 2006. Accommodating color blind computer users. In *Proc. of the 8th international ACM SIGACCESS conference on Computers and accessibility (Assets '06)*. ACM, New York, NY, USA, 40-47.

Kim, J., Vora, A. and Peli, E. 2004. MPEG based image enhancement for the visually impaired. *Optical Engineering.* 43(6): 1318-1328.

Kramer, K.M., Hedin, D.S., Rolkosky, D.J. 2010. Smartphone based face recognition tool for the

blind. In *Proc. IEEE Engineering in Medicine & Biology Society*.

Krishna, S., Little, G., Black, J., Panchanathan, S. 2005. A wearable face recognition system for individuals with visual impairments. In *Proc. of the 7th international ACM SIGACCESS conference on Computers and accessibility (Assets '05)*. ACM, New York, NY, USA, 106-113.

Kulyukin, V., Gharpure, C., Nicholson, J., and Pavithran, S. 2004. RFID in Robot-Assisted Indoor Navigation for the Visually Impaired. *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2004).*

Kurzweil, R. 2000. *The age of spiritual machines*. Penguin.

Kutiyanawala, A., Kulyukin, V., and Nicholson, J. 2011. Toward Real Time Eyes-Free Barcode Scanning on Smartphones in Video Mode . In *Proceedings of the 2011 Rehabilitation Engineering and Assistive Technology Society of North America Conference (RESNA 2011)*. Toronto, Canada.

Le Cun, Y., Boser, B., Denker, J.S. et al. 1990. Handwritten digit recognition with a back-propagation network. *Proc. Advances in Neural Information Processing Systems (NIPS).*

Lee, J. J., Lee, P. H., Koch, C. and Yuille, A. L. 2011. AdaBoost for Text Detection in Natural Scene. In *Proc. International Conference on Document Analysis and Recognition (ICDAR 2011)*.

Li, S. Z. and Jain, A. K. (eds). 2005. *Handbook of Face Recognition*. Springer.

Li, Z., Luo, G., Peli, E. 2011. Image Enhancement of high digital magnification for patients with central field loss. In *SPIE-IS&T Electronic Imaging*, SPIE Vol. 7865, Human Vision and Electronic Imaging XVI.

Liu, X. 2008. A Camera Phone Based Currency Reader for the Visually Impaired. In Proc. International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2008).

Luo, G. and Peli, E. 2011. Development and evaluation of vision rehabilitation devices. In *Proc. 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC '11)*, Boston, Massachusetts USA, August 30 - September 3, 2011. pp 5228-5231.

Manduchi, R, Coughlan J. 2012. (Computer) Vision without Sight. Communications of the ACM. 55(1).

Manduchi, R., and Kurniawan, S. 2011. Mobility-related accidents experienced by people with visual impairment. *Insight: Research and Practice in Visual Impairment and Blindness*, 4(2).

Manduchi, R., Kurniawan, S., and Bagherinia, H. 2010. Blind guidance using mobile computer vision: A usability study. *ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)*.

Mann, M. 1949. Reading machine spells out loud. *Popular Science*, February 1949.

Marston, J. R., Loomis, J. M., Klatzky, R. L., Golledge, R. G., and Smith E. L. 2006. Evaluation of spatial displays for navigation without sight. *ACM Transactions on Applied Perception* 3(2), 110–124.

Mattar, M. A., Hanson, A. R. and Learned-Miller, E. G. 2005. Sign Classification using Local and Meta-Features. In Proc. *Computer Vision Applications for the Visually Impaired (CVAVI)*, San Diego, CA.

Miele, J. 2005. User interface: The afterthought at the end of the project. Unpublished presentation at the IEEE Workshop on Computer Vision Applications for the Visually Impaired.

Molton, N., Se, S., Brady, J. M., Lee D., and Probert, P. 1998. A stereo vision-based aid for the visually impaired. *Image and Vision Computing*, 16(4).

Park, J.H. and Jeong, C.S. 2009. Real-time Signal Light Detection. International Journal of Signal Processing, Image Processing and Pattern Recognition. Vol.2, No.2.

Peli, E. and Peli, T. 1984. Image enhancement for the visually impaired. Opt. Eng. 23, 47–51.

Peli, E. 2001. Vision multiplexing - an engineering approach to vision rehabilitation device development. *Optometry and Vision Science* 78: 304-315.

Peli, E., Kim, J., Yitzhaky, Y., Goldstein, R. B. and Woods, R. L. 2004. Wide-band enhancement of television images for people with visual-impairments. *J Opt Soc Am A,* 21(6): 937-950.

Peli, E. and Woods, R. L. 2009. Image enhancement for impaired vision: the challenge of evaluation. *International Journal on Artificial Intelligence Tools,* 18(3): 415-438.

Poupyrev, I., Kato, H., and Billinghurst, M. 2000. *ARToolkit user manual.* Human Interface Technology Lab, University of Washington.

Pradeep, V., Medioni, G., and Weiland, J. 2010. Robot vision for the visually impaired. In *Proc. IEEE Workshop on Computer Vision Applications for the Visually Impaired.*

Preiser, W. and Ostroff, E. 2001. *Universal Design handbook*. McGraw-Hill Professional.

Ross, D. A. and Blasch, B. B. 2000. Wearable interfaces for orientation and wayfinding. In In Proc. ACM Conference on Assistive Technologies (ASSETS ’00).

Ross, D.A and Reynolds, M.S. 2010. RFID floors to provide indoor navigation information for people with visual impairment. *Proc. 23th Annual International Technology & Persons with Disabilities Conference (CSUN).*

Sanketi, P., Shen, H. and Coughlan, J. 2011. Localizing Blurry and Low-Resolution Text in

Natural Images. In Proc. *IEEE Workshop on Applications of Computer Vision (WACV 2011)*. Kona, Hawaii.

Sardegna, J. and Paul, T.O. *The Encyclopedia of Blindness and Visual Impairment*. Facts on File, Inc. New York. 1991.

Schoof, L.T. 1975. An analysis of Optacon usage. *American Foundation for the Blind Research Bulletin*, 29.

Se, S. 2000. Zebra-crossing Detection for the Partially Sighted. In Proc. *Conference on Computer Vision and Pattern Recognition (CVPR 2000)*. South Carolina.

Silapachote, P., Weinman, J., Hanson, A., Weiss, R. and Mattar, M. 2005. Automatic Sign Detection and Recognition in Natural Scenes. In Proc. *Computer Vision Applications for the Visually Impaired (CVAVI)*. San Diego, CA.

Stein, B.K. 1998. The Optacon: Past, present, and future. *DIGIT-EYES: The Computer Users' Network News*.

Szeliski, R. 2010. Computer Vision: Algorithms and Applications. Springer, New York.

Tang, J., Kim, J.H. and Peli, E. 2004. Image enhancement in the JPEG domain for people with vision impairment. In *IEEE Transactions on Biomedical Engineering*. 51(11): 2013-2023.

Tekin, E. and Coughlan, J. 2010. A Mobile Phone Application Enabling Visually Impaired Users to Find and Read Product Barcodes. In Proc. *International Conference on Computers Helping People with Special Needs (ICCHP '10)*. Vienna, Austria.

Tekin, E., Coughlan, J. and Shen, H. 2011. Real-Time Detection and Reading of LED/LCD Displays for Visually Impaired Persons. In Proc. *IEEE Workshop on Applications of Computer Vision (WACV 2011)*. Kona, Hawaii.

Tjan, B.S., Beckmann, P.J., Roy, R., Giudice, N., and Legge, G.E. 2005. Digital sign system for indoor wayfinding for the visually impaired. In *Proc. IEEE Workshop on Computer Vision Applications for the Visually Impaired.*

Walker, B. N. and Lindsay, J. 2005. Navigation performance in a virtual environment with bonephones. In Proc. *International Conference on Auditory Display (ICAD 2005).*

Wang, K., Babenko, B. and Belongie, S. 2011. End-to-end Scene Text Recognition. In *Proc. International Conference on Computer Vision (ICCV 2011)*. Barcelona, Spain.

Wiener, W.R., Welsh, R.L., Blasch, B.B. 2010. Foundations of Orientation and Mobility, Third Edition. AFB Press.

Winlock, T., Christiansen, E., Belongie, S. 2010. Toward real-time grocery detection for the visually impaired. In Proc. *Computer Vision Applications for the Visually Impaired (CVAVI).*,

San Francisco, CA.

Yang, X., Yuan, S. and Tian, Y. 2011. Recognizing clothes patterns for blind people by confidence margin based feature combination. In *Proc. 19th ACM International Conference on Multimedia*.

Yuan, D., and Manduchi, R. 2004. A tool for range sensing and environment discovery for the blind. In *Proc. of the IEEE Workshop on Real–Time 3D Sensor and Their Use*.