(Computer) Vision without Sight

Roberto Manduchi and
Department of Computer Engineering, University of California, Santa Cruz, Santa Cruz, CA 95064

James Coughlan
The Smith-Kettlewell Eye Research Institute, 2318 Fillmore Street, San Francisco, CA 94115

Roberto Manduchi: manduchi@soe.ucsc.edu; James Coughlan: coughlan@ski.org

Abstract

Computer vision holds great promise for helping persons with blindness or visual impairments (VI) to interpret and explore the visual world. To this end, it is worthwhile to assess the situation critically by understanding the actual needs of the VI population and which of these needs might be addressed by computer vision. This article reviews the types of assistive technology application areas that have already been developed for VI, and the possible roles that computer vision can play in facilitating these applications. We discuss how appropriate user interfaces are designed to translate the output of computer vision algorithms into information that the user can quickly and safely act upon, and how system-level characteristics affect the overall usability of an assistive technology. Finally, we conclude by highlighting a few novel and intriguing areas of application of computer vision to assistive technology.

General Terms

Algorithms; Performance; Experimentation; Human Factors

Keywords

Wayfinding; Mobility; Orientation; Guidance; Recognition

1. INTRODUCTION

More than 20 million people in the U.S. live with visual impairments ranging from difficulty seeing, even with eyeglasses, to complete blindness. Vision loss affects almost every activity of daily living. Walking, driving, reading and recognizing objects, places and people becomes difficult or impossible without vision. Technology that can assist visually impaired (VI) persons in at least some of these tasks may thus have a very relevant social impact.

Research in assistive technology for VI people has resulted in some very useful hardware and software tools in widespread use. The most successful products to date include text magnifiers and screen readers, Braille note takers, and document scanners with optical character recognition (OCR). This article focuses specifically on the use of computer vision systems and algorithms to support VI people in their daily tasks. Computer vision seems like a natural choice for these applications – in a sense, replacing the lost sense of sight with an "artificial eye." Yet, in spite of the success of computer vision technology in several other

fields (such as robot navigation, surveillance, user interface), very few computer vision systems and algorithms are currently employed to aid VI people.

In this article we review current research work in this field, analyze the causes of past failed experiences, and propose promising research directions marrying computer vision and assistive technology for the VI population. Our considerations stem in large part from our own direct experience developing technology for VI people, and from conducting the only specific workshop on "Computer Vision Applications for the Visually Impaired," which was held in 2005 (San Diego), 2008 (Marseille) and 2010 (San Francisco).

2. THE VI POPULATION

The VI community is very diverse in terms of degree of vision loss, age, and abilities. It is important to understand the various characteristics of this population if one is to design technology that is well fit to its potential "customers." Here is some statistical data, made available by American Foundation for the Blind. Of the 25 or more million Americans experiencing significant vision loss, about 1.3 million are legally blind (meaning that their visual field in their best eye is 20 degrees or less or that their acuity is less than 20/200), and only about 290,000 are totally blind (with at most some light perception). Since the needs of a low vision person and of a blind person can be very different, it is important not to over-generalize the nature of visual impairment. Another important factor to be considered is the age of a VI person. Vision impairment is often due to conditions such as diabetic retinopathy, macular degeneration and glaucoma that are prevalent at later age. Indeed, about one fourth of those reporting significant vision loss are 65 years of age or older. It is important to note that multiple disabilities in addition to vision loss are also common at later age (such as hearing impairment due to presbycusis or mobility impairment due to arthritis). Among the younger population, about 60,000 individuals in the U.S. 21 years of age or younger are legally blind. Of these, fewer than 10% use Braille as their primary reading medium.

3. APPLICATION AREAS

3.1 Mobility

In the context of assistive technology, mobility takes the meaning of "moving safely, gracefully and comfortably" [3]; it relies in large part on perceiving the properties of the immediate surroundings, and it entails avoiding obstacles, negotiating steps, drop-offs, and apertures such as doors, and maintaining a possibly rectilinear trajectory while walking. Although the population more in need of mobility aids are blind people, low-vision individuals may also occasionally trip onto unseen small obstacles or steps, especially in poor lighting conditions.

The most popular mobility tool is the white cane (known in jargon as the long cane), with about 110,000 users in the U.S. The long cane allows one to extend touch and to "preview" the lower portion of the space in front of oneself. Dog guides may also support blind mobility, but have many fewer users (only about 7,000 in the U.S.). A well trained dog guide helps maintain a direct route, recognizes and avoids obstacles and passageways that are too narrow to go through, and stops at all curbs and at the bottom and top of staircases until told to proceed. Use of a white cane or of a dog guide publicly identifies a pedestrian as blind, and carries legal obligations for nearby drivers, who are required to take special precautions to avoid injury to such a pedestrian.

A relatively large number of devices have been proposed over the past 40 years, meant to provide additional support, or possibly to replace the long cane and the dog guide altogether.

Termed Electronic Travel Aids or ETA [3], these devices typically utilize different types of range sensors (sonars, active triangulation systems, and stereo vision systems). Some ETAs are meant to simply give an indication of the presence of an obstacle at a certain distance along a given direction (clear path indicators). A number of ETAs are mounted on a long cane, thus freeing one user's hand (but at the expense of adding weight to the cane and possibly interfering with its operation). For example, the Nurion Laser Cane (no longer in production) and the Laser Long Cane produced by Vistac use three laser beams to detect (via triangulation) obstacles at head–height level, while the UltraCane (formerly BatCane) produced by Sound Foresight uses sonars on a regular cane to detect obstacles up to height level. A different type of ETA (the Sonic Pathfinder, worn as a special spectacle frame, and the Bat K–Sonar, mounted on a cane) use one or more ultrasound transducers to provide the user with something closer to a "mental image" of the scene (such as the distance and direction of an obstacle and possibly some physical characteristics of its surface.)

In recent years, a number of computer vision-based ETAs have been proposed. For example, a device developed by Yuan and Manduchi [40] utilizes structured light to measure distances to surfaces and to detect the presence of a step or a drop-off at a distance of a few meters. Step and curb detection can also be achieved via stereo vision [25]. Range data can be integrated through time using a technique called "simultaneous localization and mapping" (SLAM), allowing for the geometric reconstruction of the environment and for self-localization. Vision-based SLAM, which has been used successfully for robotic navigation, has been recently proposed as a means to support blind mobility [26, 28, 37]. Range cameras, such as the popular PrimeSense's Kinect, also represent a promising sensing modality for ETAs.

Although many different types of ETAs have appeared on the market, they have met with little success by the intended users so far. Multiple factors, including cost, usability, and performance, contribute to the lack of adoption of these devices. But the main reason is likely the fact that the long cane is difficult to surpass. The cane is economical, reliable and long-lasting, and never runs out of power. Also, it is not clear whether some of the innovative features of newly proposed ETAs (longer detection range, for example) are really useful for blind mobility. Finally, presenting complex environmental features (such as the direction and distance to multiple obstacles) through auditory or tactile channels can easily overwhelm the user, who is already concentrated on using his or her remaining sensory capacity for mobility and orientation.

Neither the long cane nor the dog guide can protect the user from all types of hazard, though. One example is given by obstacles that are at head height (such as a propped-open window or a tree branch), and thus are beyond the volume of space surveyed by the cane. In a recent survey of 300 blind and legally blind persons [21], 13% of the respondents reported that they experience head-level accidents at least once a month. The type of mobility aid (long cane or dog guide) does not seem to have a significant effect on the frequency of such accidents. Another type of hazard is represented by walking in trafficked areas, and in particular crossing a street. This requires awareness of the environment around oneself as well as of the flow of traffic, and good control of one's walking direction to avoid drift away of the crosswalk. Technology that increases the pedestrian's safety in these situations may be valuable, such as a mobile phone system using computer vision to orient the user to the crosswalk and to provide information about the timing of Walk lights [12, 13] (see Fig. 1).

3.2 Wayfinding

Orientation (or wayfinding) can be defined as the capacity to know and track one's position with respect to the environment, and to find a route to a destination. Whereas sighted persons use visual landmarks and signs in order to orient themselves, a blind person moving

in an unfamiliar environment faces a number of hurdles [20]: accessing spatial information from a distance; obtaining directional cues to distant locations; keeping track of one's orientation and location; and obtaining positive identification once a location is reached.

According to [20], there are two main ways in which a blind person can navigate with confidence in a possibly complex environment and find his or her way to a destination: piloting and path integration. Piloting means using sensory information to estimate one's position at any given time, while path integration is equivalent to the "dead reckoning" technique of incremental position estimation, used for example by pilots and mariners. Although some blind individuals excel at path integration, and can easily re-trace a path in a large environment, this is not the case for most blind (as well as sighted) persons.

Path integration using inertial sensors or visual sensors has been used extensively in robotics, and a few attempts at using this technology for blind wayfinding have been reported [18, 9]. However, the bulk of research on wayfinding has focused on piloting, with very promising results and a number of commercial products already available. For outdoor travelers, GPS represents an invaluable technology. Several companies offer GPS-based navigational systems specifically designed for VI people. None of these systems, however, can help the user in tasks such as "Find the entrance door of this building," due to the low spatial resolution of GPS reading and to the lack of such details in available GIS databases. In addition, GPS is viable only outdoors. Indoor positioning systems (for example based on multilateration from WiFi beacons) are gaining momentum, and it is expected that they will provide interesting solutions for blind wayfinding.

A different approach to wayfinding, one that doesn't require a geographical database or map, is based on recognizing (via an appropriate sensor carried by the user) specific landmarks placed at key locations. Landmarks can be active (light, radio or sound beacons) or passive (reflecting light or radio signals). Thus, rather than absolute positioning, the user is made aware of their own relative position and attitude with respect to the landmark. This may be sufficient for a number of navigational tasks, for example when the landmark is placed near a location of interest. For guidance to destinations that are beyond the landmark's "receptive field" (the area within which the landmark can be detected), a route can be built as a set of waypoints that need to be reached in sequence. Contextual information about the environment can also be provided to the VI user using digital map software and synthetic speech [14].

The best-known beaconing system for the blind is Talking Signs, now a commercial product based on technology developed at The Smith-Kettlewell Eye Research Institute[1]. Already deployed in several cities, Talking Signs uses a directional beacon of infrared light, modulated by a speech signal. This can be received at a distance of several meters by a specialized hand-held device, which also demodulates the speech signal and presents it to the user. RFID technology has also been proposed recently in the context of landmark-based wayfinding for the blind [16]. Passive RFIDs are small, inexpensive, and easy to deploy, and may contain several hundreds of bits of information. The main limitation of RFID systems is their limited reading range and lack of directionality.

A promising research direction is the use of computer vision to detect natural or artificial landmarks, and thus assist in blind wayfinding. A VI person can use their own cell phone, the camera pointing forward, to search for landmarks in view. Natural landmarks are distinctive environmental features that can be detected robustly, and used for guidance either using an existing map [11] or by matching against possibly geotagged image data sets [10,

---

[1]http://talkingsigns.com

19]. Detection is usually performed by first identifying specific keypoints in the image; the brightness or color image profile in the neighborhood of these keypoints is then represented by compact and robust descriptors. The presence of a landmark is tested by matching the set of descriptors in an image against a data set formed by exemplar images collected offline. Note that some of this research work (e.g. [11]) was aimed to support navigation in indoor spaces for persons with cognitive impairments. Apart from the display modality, the same technology is applicable for assistance to visually impaired individuals.

Artificial landmarks are meant to facilitate the detection process. For example, the color markers developed by Coughlan and Manduchi [5, 22] (see Fig. 2) are designed so as to be highly distinctive (thus minimizing the rate of false alarms) and easily detectable with very moderate computational cost (an important characteristic for mobile platforms such as cell phones with modest computing power). A similar system, designed by researchers in Gordon Legge's group at U. Minnesota, uses retro-reflective markers that are detected by a "Magic Flashlight," a portable camera paired with an infrared illuminator [33].

Artificial landmarks can be optimized for easy and fast detection by a mobile vision system. This is an advantage with respect to natural landmarks, whose robust detection is more challenging. On the other hand, artificial landmarks (as well as beacons such as Talking Signs) involve an infrastructure cost – they need to be installed and maintained, and represent an additional element to be considered in the overall environment design. This trade-off needs to be considered carefully when developing wayfinding technology. It may be argued that the additional infrastructure cost could be better justified if other communities of users in addition to the VI population would benefit from the wayfinding system. For example, even sighted individuals who are unfamiliar with a certain location (e.g. a shopping mall), and cannot read existing signs (because of a cognitive impairment, or possibly because of a foreign language barrier), may find a guidance system beneficial. Under this perspective, even the signage commonly deployed for sighted travelers can be seen as a form of artificial landmarks. Automatic reading of existing signs and, in general, of printed information via mobile computer vision, is the topic of the next section.

3.3 Printed Information Access

A common concern among the VI population is the difficulty of accessing the vast array of printed information that normally sighted persons take for granted in daily life. Such information ranges from printed documents such as books, magazines, utility bills and restaurant menus to informational signs labeling streets, addresses and businesses in out-door settings and office numbers, exits and elevators indoors. In addition, a variety of "non-document" information must also be read, including LED/LCD displays required for operating a host of electronic appliances such as microwave ovens, stoves and DVD players, and barcodes or other information labeling the contents of packaged goods such as grocery items and medicine containers.

Great progress has been made in providing solutions to this problem by harnessing OCR, which has become a mature and mainstream technology after decades of development. Early OCR systems for VI users (e.g. the Arkenstone Reader and Kurzweil Reading Machine) were bulky machines that required that the text to be read be imaged using a flatbed scanner. More recent incarnations of these systems have been implemented in portable platforms such as mobile (cell) phones (e.g. the KNFB reader[2]) and tablets (e.g. the IntelReader[3]), which allow the user to point the device's camera toward a document of interest and have it

---

[2]http://knfbreader.com
[3]http://www.intel.com/about/companyinfo/healthcare/products/reader/index.htm

read aloud in a matter of seconds. It is important to note that an important challenge of mobile OCR systems for VI users is the difficulty of aiming the camera accurately enough to capture the desired document area; thus, an important feature of the KNFB user interface is that it provides guidance to the user to help him/her frame the image properly.

However, while OCR is effective for reading printed text that is clearly resolved and which fills up most of the image, it is not equipped to find text in images that contain large amounts of unrelated clutter – such as an image of a restaurant sign captured from across the street. The problem of text detection and localization is an active area of research [4, 36, 35, 29] that addresses the challenge of swiftly and reliably sorting through visual patterns to distinguish between text and non-text patterns, despite the huge variability of text fonts and background surfaces on which they are printed (e.g. the background surface may be textured and/or curved) and the complications of highly oblique viewing perspectives, limited or poor resolution (due to large distances or motion blur) and low contrast due to poor illumination. A closely related problem is finding and recognizing signs [24], which are characterized by non-standard fonts and layouts and which may encode important information using shape (such as stop signs and signs or logos labeling business establishments).

To the best of our knowledge, there are currently no commercially available systems for automatically performing OCR in cluttered scenes for VI users. However, Blindsight Corporation's[4] "Smart Telescope" SBIR project seeks to develop a system to detect text regions in a scene and present them to a partially sighted user via a head-mounted display that zooms into the text to enable him/her to read it. Mobile phone apps such as Word Lens go beyond the functionality offered by systems targeted to VI users, such as KNFB, in that they detect and read text in cluttered scenes, though these newer systems are intended for normally sighted users.

Research is underway to expand the reach of OCR beyond standard printed text to "non-document" text such as LED and LCD displays [32], which provide access to an increasingly wide range of household appliances. Such displays pose formidable challenges that make detection and reading difficult, including contrast that is often too low (LCDs) or too high (LEDs), the prevalance of specular highlights, and the lack of contextual knowledge to disambiguate unclear characters (e.g. dictionaries are used in standard OCR to find valid words, whereas LED/LCD displays often contain arbitrary strings of digits).

Another important category of non-document text is the printed information that identifies the contents of packaged goods, which is vital when no other means of identification is available to a VI person (e.g. a can of beans and a can of soup may feel identical in terms of tactile cues). UPC bar-codes provide product information in a standardized form, and though originally designed for use with laser scanners there has been growing interest in developing computer vision algorithms for reading them from images acquired by digital cameras, especially for mobile cell platforms (e.g. the Red Laser app[5]). Such algorithms [8] have to cope with noisy and blurred images and the need to localize the bar-code in a cluttered image (e.g. taken by a VI user who has little prior knowledge of the barcode's location on a package). Some research in this area [31, 17] has specifically investigated the usability of these algorithms by VI persons, and at least one commercial system (DigitEyes[6]) has been designed specifically for the VI population. Finally, an alternative approach to package identification is to treat it as an object recognition problem ([38], see next section for

---

[4]http://www.blindsight.com/
[5]http://redlaser.com
[6]http://digit-eyes.com

details), which has the benefit of not requiring the user to locate the barcode, which comprises a small portion of the entire surface of the package.

3.4 Object Recognition

Over the past decade, increasing research efforts within the computer vision community have focused on algorithms for recognizing generic "objects" in images. For example, the PASCAL Visual Object Classes Challenge, which attracts dozens of participants every year, evaluates competing object recognition algorithms from a number of visual object classes in challenging realistic scenes[7]. Another example is Google Goggles, an online service that can be used for automatic recognition of text, artwork, book covers and more. Other commercial examples include oMoby, developed by IQ Engines, A9's SnapTell, and Microsoft's Bing Mobile application with visual scanning.

Visual object recognition for assistive technology is still in its infancy, with only a few applications proposed in recent years. For example, Winlock et al. [38] have developed a prototype system (named ShelfScanner) for assistance to a blind person while shopping at a supermarket. Images taken by a camera carried by the user are analyzed to recognize shopping items from a known set; the user is then informed about whether any of the items in his or her shopping list is in view. LookTel, a software platform for Android phones developed by IPPLEX LLC [30], performs real-time detection and recognition of different types of objects such as bank notes, packaged goods, and CD covers. The detection of doors (which can be useful for wayfinding applications) has been considered in [39].

3.5 A Human in the Loop?

The goal of the assistive technology described so far is to create the equivalent of a "sighted companion," who can assist a VI user and answer questions such as "Where am I?", "What's near me?", "What is this object?". Some researchers have begun questioning whether an automatic system is the right choice for this task. Will computer vision ever be powerful enough to produce satisfactory results in any context of usage? What about involving a "real" sighted person in the loop, perhaps through crowdsourcing? For example, the VizWiz system [2] uses Amazon's Mechanical Turk to provide a blind person with information about an object (such as the brand of a can of food). The user takes a picture of the object, which is then transmitted to Mechanical Turk's remote workforce for visual analysis, and the results are reported back to the user. The NIH-funded "Sight on Call" project by the Blindsight Corporation addresses a similar application. However, rather than relying on crowdsourcing, it uses specially trained personnel interacting remotely with the visually impaired user, on the basis of video streams and GPS data taken by the user's cell phone and transmitted to the call center.

4. INTERFACES

Each one of the systems and algorithms described above furnishes some information (e.g. the presence of an obstacle, the bearing of a landmark, or the type and brand of items on a supermarket's shelf) that needs to be presented to the VI user. This communication can use any of the user's remaining sensory channels (tactile or acoustic), but should be carefully tailored so as to provide the necessary information without annoying or tiring the user. The fact that blind persons often rely on aural cues for orientation precludes the use of regular headphones for acoustic feedback, but ear-tube earphones and bonephones [34] are promising alternatives. In the case of wayfinding, the most common methods for information display include: synthesized speech; simple audio (e.g., spatialized sound,

---

[7]Other benchmarking efforts include the TREC Video Retrieval Evaluation and the Semantic Robot Vision challenge.

generated so as it appears to come from the direction of the landmark [23]); auditory icons [6]; "haptic point interface" [23], a modality by which the user can establish the direction to a landmark by rotating a hand-held device until the sound produced has maximum volume; and tactual displays such as "tappers" [27].

One major issue to be considered in the design of an interface is whether a rich description of the scene, or only highly symbolic information, should be provided to the user. An example of the former is the vOICe, developed by Peter Mijer, which converts images taken by a live camera to binaural sound. At the opposite end are computer vision systems that "filter" incoming images to recognize specific features, and provide the user with just-in-time, minimally invasive information about the detected object, landmark or sign.

5. USABILITY

Despite the prospect of increased independence enabled by assistive technology devices and software, very few such systems have gained acceptance by the VI community as yet. We analyze in the following some of the issues that, in our opinion, should be taken into account when developing a research concept in this area. It is important to bear in mind that these usability issues can only be fully evaluated with continual feedback from the target VI population obtained by testing the assistive technology as it is developed.

5.1 Cosmetics, Cost, Convenience

No one (except perhaps for a few early adopters) wants to carry around a device that attracts unwanted attention, is bulky or inconvenient to wear or to hold, or detracts from one's attire. Often, designers and engineers seem to forget these basic tenets and propose solutions that are either inconvenient (e.g. interfering with use of the long cane or requiring a daily change of batteries) or simply unattractive (e.g. a helmet with several cameras pointing in different directions). A forward-looking extensive discussion of design for disability can be found in the beautiful book "Design Meets Disability" by G. Pullin.

Cost is also an important factor determining usability. Economics of scale is hardly achievable in assistive technology given the relatively small size of the pool of potential users, and the diversity of such a population. This typically leads to high costs for the devices that do make it to the market, which may make them unaffordable by VI users who in many cases are either retired or on disability wages.

5.2 Performance

How well should a system work before it becomes viable? The answer clearly depends on the application type. Consider for example an ETA that informs the user about the presence of a head-level obstacle. If the system produces a high rate of false alarms, the user will quickly become annoyed and turn the system off. At the same time, the system must have a very low missed detection rate, lest the user may hurt themselves against an undetected obstacle, possibly resulting in medical (and legal) consequences. Other applications may have less stringent requirements. For example, in the case of a cell phone-based system that helps one find a certain item in the grocery store, no harm will be caused to the user if the item is not found or if the wrong item is selected. Still, poor performance is likely to lead to users abandoning the system. Establishing functional performance metrics and assessing minimum performance requirements for assistive technology systems is still an open and highly needed research topic.

5.3 Mobile Vision and Usability

The use of mobile computer vision for assistive technology imposes particular functional constraints. Computer vision requires use of one or more cameras to acquire snapshots or video streams of the scene. In some cases, the camera may be hand-held, for example when embedded in a cell phone. In other cases, a miniaturized camera may be worn by the user, perhaps attached to one's jacket lapel or embedded in one's eyeglasses frames. The camera's limited field of view is an important factor in the way the user interacts with the system to explore the surrounding environment: if the camera is not pointed towards a feature of interest, this feature is simply not visible. Thus, it is important to study how a visually impaired individual, who cannot use feedback from the camera's viewfinder, can maneuver the camera in order to explore the environment effectively. Of course, the camera's field of view could be expanded, but this typically comes at the cost of a lower angular resolution. Another possibility, explored by Winlock et al. [38], is to build a panoramic image by stitching together several images taken by pointing the camera in different directions.

It should be noted that, depending on the camera's shutter speed (itself determined by the amount of light in the scene), pictures taken by a moving camera may be blurred and difficult or impossible to decipher. Thus, the speed at which the user moves the camera affects recognition. Another important issue is the effective frame rate, that is, the number of frames per second that can be processed by the system. If the effective frame is too low, visual features in the environment may be missed if the user moves the camera too fast in the search process. For complex image analysis tasks, images can be sent to a remote server for processing (e.g. the LookTel platform [30]), in which case the speed and latency are determined by the communication channel. Hybrid local/remote processing approaches, with scene or object recognition performed on a remote sever and fast visual tracking of the detected feature performed by the cell phone, may represent an attractive solution for efficient visual exploration.

Thus, a mobile vision system for assistive technology is characterized by the interplay between camera characteristics (field of view, resolution), computational speed (effective achievable frame rate for a given recognition task), and user interaction (including the motion pattern used to explore the scene, possibly guided by acoustic or tactile feedback). Preliminary research work has explored the usability of such systems for tasks such as wayfinding [22] and access to information embedded in bar codes [31, 17].

6. CONCLUSIONS AND NEW FRONTIERS

Advances in mobile computer vision hold great promise for assistive technology. If we can teach computers to see, they may become a valuable support for those of us whose sight is compromised or lost. However, decades-long experience has shown that creating successful assistive technology is difficult. Far too often, engineers have proposed technology-driven solutions that either do not directly address the actual problems experienced by VI persons, or that are not satisfactory in terms of performance level, ease of use, or convenience. Assistive technology is a prime example of user-centered technology: the needs, characteristics, and expectations of the target population must be understood and taken into account throughout the project, and must drive all of the design choices, lest the final product result in disappointment for the intended user, and frustration for the designer. Our hope is that a new generation of computer vision researchers will take on the challenge, and arm themselves with enough creativity to produce innovative solutions, and humbleness to listen to the persons who will use this technology.

In closing this contribution, we would like to propose a few novel and intriguing application areas that in our opinion deserve further investigation by the research community.

6.1 Independent Wheeled Mobility

One dreaded consequence of progressive vision loss (for example, due to an age-related condition) is the ensuing loss of driving privileges. For many individuals, this is felt as a severe blow to their independence. Alternative means of personal wheeled mobility that do not require a driving license could be very desirable to active individuals who still have some degree of vision left. For example, some low-vision persons reported good experience using the two-wheel Segway, driven on bicycle lanes [1]. These vehicles could be equipped with range and vision sensors to improve safety, minimizing the risk of collisions and ensuring that the vehicle remains within a marked lane. With the recent emphasis on sensors and machine intelligence for autonomous cars in urban environments, it is only reasonable that the VI community should soon benefit from these technological advances.

6.2 Blind Photography

Many people find it surprising that people with low vision or blindness enjoy photography as a recreational activity. In fact, a growing community of VI photographers take and share photos of family and friends, of objects, and of locations they have visited; some have elevated the practice of photography to an art form, transforming what would normally be considered a challenge (the visual impairment) into an opportunity for creativity. There are numerous websites (e.g. http://blindwithcameraschool.org), books and art exhibitions focused on this subject, which could present an interesting opportunity for computer vision researchers. A variety of computer vision techniques such as face detection, geometric scene analysis and object recognition could help a VI user correctly orient the camera and frame the picture. Such techniques, when coupled with a suitable interface, could provide a VI person with a feedback mechanism similar to the viewfinder used by sighted photographers.

6.3 Social Interaction

Blindness may, among other things, affect one's interpersonal communication skills, especially in scenarios with multiple persons interacting (e.g. in a meeting). This is because communication in these situations is largely non-verbal, relying on cues such as facial expressions, gaze direction, and other forms of the so-called "body language." Blind individuals cannot access these non-verbal cues, leading to a perceived disadvantage that may result in social isolation. Mobile computer vision technology may be used to capture and interpret visual cues from other persons nearby, thus empowering the VI user to participate more actively in the conversation. The same technology may also help a VI person become aware of how he or she is perceived by others. A survey conducted with 25 visually impaired persons and 2 sighted specialists [15] has highlighted some of the functionalities that would be most desirable in such a system. These include: understanding whether one's personal mannerisms may interfere with social interactions with others; recognizing the facial expressions of other interlocutors; and knowing the names of the people nearby.

6.4 Assisted Videoscripting

Due to their overwhelmingly visual content, movies are usually considered inaccessible to blind people. In fact, a VI person may still enjoy a movie from its soundtrack, especially in the company of friends or family. In many cases, though, it is difficult to correctly interpret ongoing activities in the movie (for example, where the action is taking place, which characters are currently in the scene and what they are doing) from the dialogue alone. In addition, many relevant non-verbal cues (such as the facial expression of the actors) are lost.

Videodescription (VD) is a technique meant to increase accessibility of existing movies to VI persons by adding a narration of key visual elements, which is presented to the listener during pauses in the dialogue. Although the VD industry is fast growing, due to increasing demand, the VD generation process is still tedious and time-consuming. This process, however, could be facilitated by the use of semi-automated visual recognition techniques, which have been developed in different contexts (such as surveillance and video database indexing). An early example is VDManager [7], a VD editing software tool, which uses speech recognition as well as key-places and key-faces visual recognition.

Acknowledgments

References

1. Ackel W. A Segway to independence. Braille Monitor. 2006

2. Bigham, JP.; Jayant, C.; Ji, H.; Little, G.; Miller, A.; Miller, RC.; Miller, R.; Tatarowicz, A.; White, B.; White, S.; Yeh, T. VizWiz: Nearly real-time answers to visual questions. Proc. ACM Symposium on User Interface Software and Technology, UIST '10; 2010.

3. Blasch, B.; Wiener, W.; Welsh, R. Foundations of Orientation and Mobility. 2. AFB Press; 1997.

4. Chen, X.; Yuille, A. Detecting and reading text in natural scenes. Proc. IEEE Conference on Computer Vision and Pattern Recognition, CVPR '04; 2004.

5. Coughlan J, Manduchi R. Functional assessment of a camera phone-based wayfinding system operated by blind and visually impaired users. International Journal on Artificial Intelligence Tool. 2009; 18(3):379–397.

6. Dingler, T.; Lindsay, J.; Walker, BN. Learnability of sound cues for environmental features: Auditory icons, earcons, spearcons, and speec. Proc. International Conference on Auditory Display (ICAD 2008); 2008.

7. Gagnon, L.; Chapdelaine, C.; Byrns, D.; Foucher, S.; Heritier, M.; Gupta, V. A computer-vision-assisted system for Videodescription scripting. Proc. Workshop on Computer Vision Applications for the Visually Impaired, CVAVI '10; 2010.

8. Gallo O, Manduchi R. Reading 1-D barcodes with mobile phones using deformable templates barcodes with mobile phones using deformable templates. IEEE Transactions on Pattern Analysis and Machine Intelligence. in press.

9. Hesch JA, Roumeliotis SI. Design and analysis of a portable indoor localization aid for the visually impaired. International Journal on Robotics Research. Sep.2010 29:1400–1415.

10. Hile H, Liu A, Borriello G, Grzeszczuk R, Vedantham R, Kosecka J. Visual navigation for mobile devices. IEEE Multimedia. 2010; 17(2):16–25.

11. Hile, H.; Vedantham, R.; Cuellar, G.; Liu, A.; Gelfand, N.; Grzeszczuk, R.; Borriello, G. Landmark-based pedestrian navigation from collections of geotagged photos. Proc. International Conference on Mobile and Ubiquitous Multimedia, MUM '08; 2008.

12. Ivanchenko, V.; Coughlan, J.; Shen, H. Crosswatch: A camera phone system for orienting visually impaired pedestrians at traffic intersections. Proc. International Conference on Computers Helping People with Special Needs, ICCHP '08; 2008.

13. Ivanchenko, V.; Coughlan, J.; Shen, H. Real-time walk light detection with a mobile phone. Proc. International Conference on Computers helping people with special needs, ICCHP '10; 2010.

14. Kalia AA, Legge GE, Ogale A, Roy R. Assessment of indoor route-finding technology for people who are visually impaired. Journal of Visual Impairment & Blindness. Mar; 2010 104(3):135–147. [PubMed: 21869851]

15. Krishna, S.; Colbry, D.; Black, J.; Balasubramanian, V.; Panchanathan, S. A systematic requirements analysis and development of an assistive device to enhance the social interaction of

people who are blind or visually impaired. Proc. Workshop on Computer Vision Applications for the Visually Impaired, CVAVI '08; 2008.

16. Kulyukin V, Kutiyanawala A. Accessible shopping systems for blind and visually impaired individuals: Design requirements and the state of the art. The Open Rehabilitation Journal. 2010; 2

17. Kutiyanawala, A.; Kulyukin, V. An eyes-free vision-based UPC and MSI barcode localization and decoding algorithm for mobile phones. Proc. Envision Conference; 2010.

18. Ladetto, Q.; Merminod, B. Combining gyroscopes, magnetic compass and GPS for pedestrian navigation. Proc. Int. Symposium on Kinematic Systems in Geodesy, Geomatics and Navigation, KIS '01; 2001.

19. Liu, J.; Phillips, C.; Daniilidis, K. Video-based localization without 3D mapping for the visually impaired. Proc. Workshop on Computer Vision Applications for the Visually Impaired, CVAVI '10; 2010.

20. Loomis, JM.; Golledge, RG.; Klatzky, RL.; Marston, JR. Assisting wayfinding in visually impaired travelers. In: AG, editor. Applied Spatial Cognition: From Research to Cognitive Technology. Lawrence Erlbaum Assoc; Mahwah, NJ: 2007. p. 179-202.

21. Manduchi R, Kurniawan S. Mobility-related accidents experienced by people with visual impairment. AER Journal: Research and Practice in Visual Impairment and Blindness. in press.

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

23. Marston JR, Loomis JM, Klatzky RL, Golledge RG, Smith EL. Evaluation of spatial displays for navigation without sight. ACM Transactions on Applied Perception. 2006; 3(2):110–124.

24. Mattar, MA.; Hanson, AR.; Learned-Miller, EG. Sign classification using local and meta-features. Proc. Workshop on Computer Vision Applications for the Visually Impaired, CVAVI '05; 2005.

25. Pradeep, V.; Medioni, G.; Weiland, J. Piecewise planar modeling for step detection using stereo vision. Proc. Workshop on Computer Vision Applications for the Visually Impaired, CVAVI '08; 2008.

26. Pradeep, V.; Medioni, G.; Weiland, J. Robot vision for the visually impaired. Proc. Workshop on Applications of Computer Vision for the Visually Impaired, CVAVI '10; 2010.

27. Ross, DA.; Blasch, BB. Wearable interfaces for orientation and wayfinding. Proc. ACM SIGACCESS Conference on Computers and Accessibility, ASSETS '00; 2000.

28. Saez, J.; Escolano, F. Stereo-based aerial obstacle detection for the visually impaired. Proc. Workshop on Computer Vision Applications for the Visually Impaired, CVAVI '08; 2008.

29. Sanketi, P.; Shen, H.; Coughlan, J. Localizing blurry and low-resolution text in natural images. Proc. IEEE Workshop on Applications of Computer Vision, WACV '11; 2011.

30. Sudol, J.; Dialameh, O.; Blanchard, C.; Dorcey, T. Looktel: A comprehensive platform for computer-aided visual assistance. Proc. Workshop on Computer Vision Applications for the Visually Impaired, CVAVI '10; 2010.

31. Tekin, E.; Coughlan, J. An algorithm enabling blind users to find and read barcodes. Proc. IEEE Workshop on Applications of Computer Vision, WACV '09; 2009.

32. Tekin, E.; Coughlan, J.; Shen, H. Real-time detection and reading of LED/LCD displays for visually impaired persons. Proc. IEEE Workshop on Applications of Computer Vision, WACV '11; 2011.

33. Tjan, BS.; Beckmann, PJ.; Roy, R.; Giudice, N.; Legge, GE. Digital sign system for indoor wayfinding for the visually impaired. Proc. Workshop on Computer Vision for the Visually Impaired, CVAVI '05; 2005.

34. Walker, BN.; Lindsay, J. Navigation performance in a virtual environment with bonephones. Proc. International Conference on Auditory Display (ICAD2005); 2005. p. 260-3.

35. Wang, K.; Belongie, S. Word spotting in the wild. Proc. European Conference on Computer Vision (ECCV); 2010.

36. Weinman JJ, Learned-Miller E, Hanson AR. Scene text recognition using similarity and a lexicon with sparse belief propagation. IEEE Transactions on Pattern Analisis and Machine Intelligence. Oct.2009 31:1733–1746.

37. Wilson, J.; Walker, BN.; Lindsay, J.; Cambias, C.; Dellaert, F. SWAN: System for wearable audio navigation. Proc. IEEE International Symposium on Wearable Computers; 2007.

38. Winlock, T.; Christiansen, E.; Belongie, S. Toward real-time grocery detection for the visually impaired. Proc. Workshop on Computer Vision Applications for the Visually Impaired, CVAVI '10; 2010.

39. Yang, X.; Tian, Y. Robust door detection in unfamiliar environments by combining edge and corner features. Proc. Workshop on Computer Vision Applications for the Visually Impaired, CVAVI '10; 2010.

40. Yuan, D.; Manduchi, R. Dynamic environment exploration using a virtual white cane. Proc. IEEE Conference on Computer Vision and Pattern Recognition, CVPR '05; 2005.

Figure 1.
Crosswatch system for providing guidance to VI pedestrians at traffic intersections. (a) Blind user "scans" the crosswalk by panning cell phone camera left and right, and system provides feedback to help user align him/herself to crosswalk before entering it. (b) Schematic shows that system announces to user when the Walk light is illuminated.

Figure 2.
Experiments with a blind user searching for a landmark (represented by a color marker placed on the wall) using a camera cell phone (from [22]).