A new model based on task recognition and monitoring for the development of sensory substitution assistive systems for the visually impaired

Pantelis Elinas, Yi Li, Lochana Perera
Australian Centre for Field Robotics
The University of Sydney
Sydney, NSW 2006

Abstract

According to the World Health Organization more than 314 million people worldwide suffer from some form of visual impairment with 87% of them living in the developing world. Clearly, there exists a need for the development of cost-effective assistive technologies for improving the quality of life for the visually impaired. Sensory substitution systems aim to replace one sensory signal, e.g., vision, with another, e.g., touch, delivered via vibrotactile or electrotactile stimulation. Although much progress has been made towards the development of such systems some of which have been made available commercially their capabilities and usability are still far from desired. In this paper, we survey recent advances in sensory substitution with a focus on developing assistive devices for the visually impaired, identify a number of roadblocks to the development of more advanced systems and propose a new development model utilizing state-of-the-art machine learning techniques. We believe that the proposed model will allow us to lift some of the obstacles preventing the mass adaptation of such assistive devices.

1 Introduction

In this paper, we review the state-of-the-art in sensory substitution systems for the visually impaired (SSSVI) with an emphasis on Tactile-Vision sensory substitution (TVSS), identify the limitations of the development model behind these systems and propose a new model based on the latest advancements in machine learning in an effort to remove some of the roadblocks preventing SSSVIs from becoming more widely used.

Visual impairment of various levels from partial to total loss of vision is widespread in the global population. According to the World Health Organization (WHO) there are 314 million people suffering from visual impairments worldwide with 45 million of these classified as totally blind [2]. According to a 2008 report by Prevent Blindness America, in the USA, there are more than 3.6 million adults over 40 years of age suffering from different levels of visual impairment including total loss of vision [30]; the Braille Institute reports 21% of adults over the age of 65 have some kind of visual impairment. A 2007 study by Prevent Blindness America focused on the economics of adult blindness in the U.S. states that the total cost for treating visual impairment reaches 51.4 billion dollars annually [29]. In a similar study, the economic impact of blindness in Australia alone was estimated for the year 2004 to be 9.85 billion dollars [1].

In addition, the WHO reports that the majority of people suffering a loss of visual acuity live in the developing world. It is a staggering 87% of the total [2]. This means that any technology designed to aid the visually impaired must be not only effective but also affordable.
Clearly, there is a need for technologies to tackle visual impairment in all its forms both in terms of improving the quality of life of a large number of people but also because of an economic incentive. Moreover, because of the large need for such technologies in the developing world, keeping their cost low is of the utmost importance.

Today, the visually impaired benefit from the use of a number of assistive technologies. Those suffering from total loss of vision most commonly utilize assistive technologies such as the Braille system for reading and the long cane for mobility. Guide dogs are also in use as mobility aids but they are not as common as the long cane. Magnifying technologies for those with only partial loss of vision are also of widespread use. Recently, researchers have attempted to improve these traditional aid devices; for example, researchers have created an improved long cane capable of receiving information from a grid of RFID tags embedded in the environment coupled with an off-the-shelf orientation and navigation system [10]. A similar system that uses a laser sensor for localization was recently proposed [14]. Others have studied the possibility of using robots as a replacement for guide dogs [21, 22]. Our focus is on new technologies for assisted living based on the principle of sensory substitution.

Sensory substitution is the process by which sensory stimuli from one sense, e.g., vision, can be adequately replaced with stimuli from another sense, e.g., touch. Coupled with encouraging results from brain plasticity research [23] showing how the brain can adapt such that neurons previously used to process data from one sense can be recruited to enhance a person’s ability to process data from another sense and as such used to substitute one sense for another, researchers have developed new systems for assistive living with much effort focused on helping the visually impaired. Paul Bach-y-Rita pioneered this line of work starting in the 1960s resulting in a number of commercially available systems with, unfortunately, still limited capabilities. In the next section, we describe a number of sensory substitution systems designed to improve the quality of life for the visually impaired.

The rest of this paper is structured as follows. Section 2 provides an overview of the most recent and successful commercial and research SSSVIs. In Section 3 we present the common design behind current SSSVIs and identify the main bottleneck in developing effective systems. In Section 4, we propose a new model for developing SSSVIs in an effort to overcome some of the limitations inherent in the current model. We conclude in Section 5.

2 Related work

There has been much interest in sensory substitution applied to the design and development of assistive devices for the visually impaired. Many different systems have been proposed that transform images captured using one or more cameras into either audio or haptic feedback. The main idea behind such systems is to exploit the brain’s ability to adapt to a loss of a sense by utilizing the excess neurons to process data from another, still functioning sense. In this section we provide an overview of the most significant and recent developments in SSSVIs. For a more comprehensive overview of the many different systems that have been studied in the past, the reader is directed to one or both of [5, 31].

In auditory-vision sensory substitution, a camera is used to capture images and a computer to process them and translate them into an audio signal allowing the user to navigate an obstacle course or recognize and distinguish among a small number of objects, e.g., signs.

One of the earliest auditory-vision sensory substitution systems was “The Voice” [28] which translated a grey scale image of size $64 \times 64$ pixels to a sinusoidal audio wave modulating its frequency according to the vertical position of the detected features and its amplitude according to their brightness. Scanning the images from left to right, it would take 640ms to map a single image to sound.

Recent advances in computer vision [12, 26] allow the recognition of objects within an image with a high degree of accuracy and researchers have exploited this to develop new assistive systems. For example, [8] presents a system that uses a state-of-the-art object recognition system to identify objects under difficult viewing conditions and provide 2 kinds of auditory feedback: one is to speak the name of the object identified and the other is a beacon that allows the user to determine the direction to the object. The user interacts with the system using voice input. The researchers report
that a blind person utilizing this system could localize a sign using the beacon interface in less than 12 seconds on average.

See CoLoR is another similar system specifically designed for navigational purposes [6]. The main idea behind it is to map image colors to musical instrument sounds. Distances to objects are encoded by varying the duration of the sound among 4 different levels.

One of the main problems for auditory-vision sensory substitution is that it blocks a sense that is often needed for verbal communication with other people or for perceiving natural environmental sounds that may be important to the end user. For this reason, we now turn to tactile-vision sensory substitution (TVSS) systems.

Tactile-vision sensory substitution systems map image data to haptic stimuli using, most commonly, either a vibrotactile or electrotactile human-machine interface (HMI). The latter use electrodoses to supply small electrical signals to excite the tactile sensors in a person’s skin [18]. One of the earliest such systems that has found commercial use in recent years is the TVSS presented in [35, 9]. For this system, the electrotactile stimuli are supplied to the user’s tongue because it has been found to be one of the most sensitive areas on the body that can be stimulated and the most information transmitted to the brain. Recently, researchers have also studied carefully the characteristics of electrotactile stimulation on a person’s fingertips [34]. The Electro-Neural Vision System places an electrode on each finger and maps data from a stereo camera to appropriate electrical signals meant to help a blind person navigate a complex environment [27].

One of the major disadvantages of electrotactile stimulation is that continuous use can irritate and possibly damage the part of the skin stimulated. Vibrotactile stimulation, i.e., stimulation using small vibrating motors, has been studied as an alternative along with other methods for haptic feedback utilizing less portable force-feedback devices such as the PHANToM [33] or others like it [32]. Many systems have been developed over time with varying degrees of success none of which has been commercially successful [5]. More recently, researchers have focused on the use of small vibrating motors primarily used in mobile devices as the preferred method for haptic stimulation. In [36], an affordable haptic glove that draws power from a computer’s USB port is developed for use in navigation. In [13], an array of small vibrating motors are attached to a user’s back and coupled with a stereo vision system used to, once more, aid in navigation. Similarly, in [17], a system to aid in mobility maps data from depth images captured using a stereo camera to haptic stimuli using an array of vibrating motors placed on a belt around the user’s waist. Finally, [15] developed a similar haptic belt for the same purpose but instead of stereo vision they use GPS information from a PDA for localization. They tested their system with a small sample of volunteers with visual impairments ranging from total to partial blindness and found that for the navigational task haptic feedback is as effective as other more traditional aids such as the commonly used long cane.

3 Current SSSVI design and its limitations

In the previous section we outlined a number of sensory substitution assistive systems for those suffering from visual impairment. Most of these systems follow a common design as shown in Figure 1. In this model, image data captured using a camera are processed for feature extraction, object recognition, and/or stereo vision computation and then mapped to vibrotactile or electrotactile haptic stimuli. This mapping is decided beforehand and it is always task specific.

Using the same basic model we have recently built a vibrotactile sensory substitution system the components of which are shown in Figure 2 and the full setup is shown in Figure 3. We use small, 10mm vibrating motors made by Precision Microdrives designed for use in mobile devices such as cell phones. For control, we use a standard micro-controller, the ATmega 2560 with enough outputs to drive 6 motors. A serial communication (RS232) and Bluetooth setup allows us to interface the micro-controller with a laptop computer providing additional computer power for image processing.

Using this setup, we ran a small, preliminary user study involving 10 subjects [24]; our goal was to determine their ability to distinguish between two stimuli applied consecutively on the finger with a small delay inbetween. We found that a stimulus between 200ms and 500ms in duration was comfortable and easy for the subjects to detect; two consecutive pulses had to be separated by 500ms for the subjects to tell the difference between them. All study participants were males.
between the ages of 24 and 40; some of them suffered from some visual impairment requiring the use of corrective glasses but none was blind.

One of the major problems with the basic model is the need to map data from a source, e.g., images, with high information content (the estimated bandwidth of the human visual system is 8.75 kilobits per second [20]) to one with limited bandwidth, i.e., haptic. Researchers have found that the perceptual information rate for a vibrotactile system is just $2 - 56$ bits per second [19]. As a result, systems designed using the current model apply heuristic methods to limit the information passed to the user to what is considered the most relevant to communicate with respect to the user’s task.

So, the bandwidth bottleneck is one of the main problems that must be overcome before we can develop a widely useful TVSS. However, there is a second problem that prevents the wider adaptation of such devices. In Section 2, we reviewed a number of SSSVIs with an emphasis on those designed to aid in navigation. It has been demonstrated by researchers that such task-specific SSSVIs can be effective with some commercial applications, e.g., wayfinding. However, there is a fundamental problem with these systems which at the same time is their major strength: they are task specific. An
SSSVI specifically developed for navigation is good for this and only this task. To better understand why such task specific SSSVIs are insufficient, consider the following use case.

A blind person is equipped with a haptic glove specifically designed for wayfinding such as the system presented in [17]. The person receives haptic feedback depending on the distance to obstacles around her. Let us assume that this person is walking down a long corridor on her way out of a building. Half way down the corridor, she comes across a friend who stops in front of her to chat. At this stage, the vibrotactile system would continuously indicate the presence of an obstacle dead ahead. This task-specific system continues to operate to the point of annoyance unless explicitly turned off. A more proper system should recognize that the user has changed task and is now interacting with another person and either automatically turn off or provide feedback according to the current situation.

4 Proposed model for SSSVI design

In the previous section we gave an overview of the common model employed in the development of sensory substitution assistive systems for the visually impaired. At the same time, we identified two major limitations of the current model: one is an information bottleneck caused by the limited bandwidth of the tactile sense constraining the effectiveness of such devices; second is the task-specific design of the devices that limit their usefulness. In this section, we propose an extension to the current model that can potentially remove these two obstacles.

The new model that we propose for the design of SSSVIs addresses the above two issues by combining intelligent modules for task recognition and monitoring. It is illustrated diagrammatically in Figure 4.

Clearly, both task switching and monitoring are non-trivial problems to solve but both are well studied and some current solutions can be transferred to the domain of sensory substitution for assistive devices. Task switching requires task recognition which has been studied extensively in several domains at different levels of abstraction [11, 25]. In [4], it is shown that activity recognition from biometric data is possible at a high accuracy of over 80%. In [25], GPS data are used within a Hierarchical Conditional Random Field for the recognition of a person’s high-level activity. It is shown experimentally that the approach is capable of segmenting a person’s daily activities to one of several categories including working, traveling, visiting, etc.
At the same time, task monitoring has been demonstrated as an effective way for developing assistive technologies for helping with a hand washing task for elderly people with dementia [16]. In this latter project, a Partially Observable Markov Decision Process [7] is used to monitor the user’s progress in the hand washing task and provide auditory feedback accordingly. A similar model could be used to develop an SSSVI which provides task-specific feedback in a principled as opposed to hand-coded way. We expect that employing such an approach will allow us to better utilize the limited bandwidth of our vibrotactile sensory substitution system.

5 Conclusions

In this paper, we have considered the problem of building sensory substitution devices for helping those suffering from visual impairment and we have reviewed some recent developments in sensory substitution as applied to building such assistive living devices. We presented our own cost-effective, vibrotactile sensory substitution apparatus and the preliminary results from a small user study that we conducted.

Most importantly, we have identified two main shortcomings with the current SSSVI design model that limits their usefulness, i.e., limited bandwidth of the tactile sense and the single task design of the devices. Finally, we proposed a new model that extends the traditional approach with intelligent task recognition and monitoring potentially eliminating the information bottleneck and limited task applicability of current systems.

As we said earlier, the vast majority of people who suffer from visual impairment live in the developing world and as such any assistive technology developed must be low cost. The non-invasive, vibrotactile sensory substitution system we presented in Section 3 can be built for only a few hundred dollars the highest cost being the microcontroller as the motors can be had for just a couple of dollars each. External processing power for visual processing and task recognition and monitoring could potentially be provided by low cost computing hardware found in modern PDAs and smartphones.

Acknowledgments

This work is supported by the Rio Tinto Centre for Mine Automation and the ARC Centre of Excellence programme funded by the Australian Research Council (ARC) and the New South Wales State Government.
References


