

Received 8 September 2013; revised 11 January 2014; accepted 8 February 2014; Date of publication 6 March 2014;  
date of current version 19 March 2014.

Digital Object Identifier 10.1109/JTEHM.2014.2309343

# FingerSight: Fingertip Haptic Sensing of the Visual Environment

SAMANTHA HORVATH<sup>1</sup>, JOHN GALEOTTI<sup>1</sup>, BING WU<sup>2</sup>, ROBERTA KLATZKY<sup>3</sup>,  
MEL SIEGEL<sup>1</sup> (Fellow, IEEE), AND GEORGE STETTEN<sup>4</sup> (Member, IEEE)

<sup>1</sup>Robotics Institute, Carnegie Mellon University, Pittsburgh, PA 15213, USA

<sup>2</sup>College of Technology and Innovation, Arizona State University, Mesa, AZ 85212, USA

<sup>3</sup>Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA

<sup>4</sup>Department of Bioengineering, University of Pittsburgh, PA 15261, USA

CORRESPONDING AUTHOR: G. STETTEN (george@stetten.com)

This work was supported by NIH under Grant 1R01EY021641 and by NSF GRFP 0946825.

**ABSTRACT** We present a novel device mounted on the fingertip for acquiring and transmitting visual information through haptic channels. In contrast to previous systems in which the user interrogates an intermediate representation of visual information, such as a tactile display representing a camera generated image, our device uses a fingertip-mounted camera and haptic stimulator to allow the user to feel visual features directly from the environment. Visual features ranging from simple intensity or oriented edges to more complex information identified automatically about objects in the environment may be translated in this manner into haptic stimulation of the finger. Experiments using an initial prototype to trace a continuous straight edge have quantified the user's ability to discriminate the angle of the edge, a potentially useful feature for higher levels analysis of the visual scene.

**INDEX TERMS** Visually impaired, blind, haptics, sensory substitution.

## I. INTRODUCTION

The visual environment provides a vast assortment of information to the sighted individual during the acts of daily living. For the visually-impaired individual, a need exists to compensate for the lack of this information, which has motivated the introduction of a wide variety of devices that transfer at least some visual information to another sense. However, the primary assistive technologies widely used today by the visually-impaired to navigate through the environment are essentially unchanged from those of twenty years ago, namely, white canes and guide dogs [1]. Although these two methods can facilitate the ability to travel safely in a variety of indoor and outdoor environments, neither provide the kind of assistance needed to straighten a picture frame on the wall or find a can of soup on a counter-top. Electronic navigation aids are finding some acceptance, for example, ultrasonic canes (UltraCane from Sound Foresight Technology, Ltd.) that provide tactile cues to objects beyond the tip of the cane, as well as portable computers with global positioning systems (GPS) and electronic Braille or speech interfaces [2]. However, replacing the more general capabilities of vision to

provide detailed information about objects in the environment has proven more difficult.

There exists a need for a device that allows active interrogation and sensing of the 3D visual environment surrounding the operator while moving through everyday environments. There also exists a need for a device that not only allows the user to sense the environment but also provides control of specific aspects of the environment that have been detected. These abilities to interrogate and control the environment should not be limited to a specific, predetermined environment, such as those that already contain infrared (IR) transmitters [3]. Devices using a concept we call FingerSight are intended to allow visually impaired users to remotely identify objects of interest in, and navigate through, natural environments, by acting as a sensory substitution system for the user's sense of sight. There are already a number of technologies that monitor hand, eye, and body motion from a fixed or mobile camera to interpret gestures, commands, gaze direction, e.g. [4]. Some of these may be used to track the operator's interrogation of the environment, as with the commercially available eye-tracking glasses (e.g., Tobii, Tokyo),

but more often they are used to interpret the user's motions as commands.

Our original goal was, and largely remains, to serve the population who are visually impaired. Most devices designed with this goal make use of sensory substitution systems, which transmit stimuli normally interpreted by one sense through another sense. Sensory substitution systems generally consist of three components: a sensor that collects a signal, a coupling system that processes the signal, and finally an output transducer, which transmits the signal to a sense organ not normally used for that signal [6], [7]. The coupling system may also perform some analysis/interpretation of the signal. We specifically consider substitution for the visual system by the haptic system (i.e. one's sense of touch) [8]. Although hearing offers a potentially higher bandwidth than touch, it is crucial not to impede the existing use of hearing, which can be acutely well developed by those who lack normal vision, providing essential acoustic cues about the environment [5]. As for touch, the hands offer the greatest versatility and sensitivity, but still we must not completely usurp their use, since they are essential for so many tasks in daily living.

Various vision-to-touch sensory substitution systems have been devised. Some employ tactile pin arrays, which feature a grid of small pins with adjustable heights. Tactile pin arrays can be used for either passive stimulation of the fingers or during active exploration to simulate contact with surfaces and objects. Existing systems capture an image and then display that image on a tactile screen that can be worn on a belt, for example. The fingers can then be used to interrogate the image depicted on the tactile screen in an effort to "visualize" the image [9], [10]. Some devices have been developed to directly convey a predetermined map from coordinates of physical space to a tactile spatial layout. An example is Bach-y-Rita's Brainport device, which uses an array of electrodes placed on the tongue to relay camera information [11]. Sensory substitution systems also exist that provide an audio output, such as text-to-speech devices that employ optical character recognition, canes that provide audio feedback to enhance the natural tactile forces detected at the tip itself [12], and the vOICe system, which maps image pixels into frequency and time [13].

Other devices use vibrotactile-based systems, especially to elicit sensations of texture. Vibrotactile sensitivity is found in mechanoreceptors lying within the subcutaneous tissue, which respond maximally to frequencies of 250-300 Hz, as well as kinesthetic receptors [14]. The CyberTouch™ Glove (Immersion Corp., San Jose, CA) uses optical tracking of each finger to determine when vibratory stimulators on that finger should be activated. Some sensory aids, such as the Optacon (Telesensory Corp., defunct), aim to simulate tactile exploration of virtual objects and have been developed specifically for reading the printed page [15]. Our proposed device uses a vibrotactile stimulator to provide visual information gathered from a color camera system. Both the camera and stimulator are mounted on the user's finger. The information from the

camera system can be processed into a single feature whose presence or absence is communicated through the stimulator. For example, the camera may provide a positive signal when an edge is detected in the field of view.

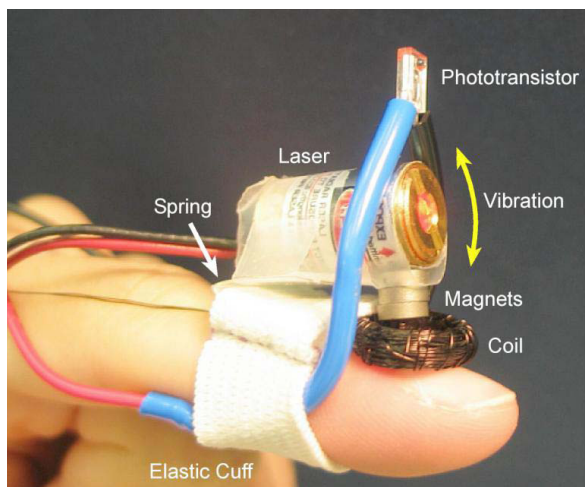
Our system is similar in some ways to that of Lenay [6], [16] who attached a photocell to the tip of a finger and used it to activate a vibrator held in the other hand in an all-or-nothing manner. The subject was capable of locating light sources in a 3D space. Another system, developed by Burch and Pawluk, is mounted on the same finger as the sensor, for interacting with graphical objects on a flat screen [17]. This system senses a single RGB pixel at a short distance (i.e., to the screen that the subject is touching), and vibrates based upon the color sensed. Our system goes beyond this basic idea by detecting not just the general directional presence of light or contact with a single light source, but rather actual images (or a series of images over time) using a miniature camera. Like Burch and Pawluk, we put the vibrator on the same finger as the detector. We believe that this placement promotes an intuitive coupling between the scene pointed to by the camera mounted on the fingertip and the feedback felt by that finger.

In general, it is imperative for designers of any sensory substitution system to consider not only what is technologically feasible, but also what is functional in the context of the sensory and cognitive limitations of the user. In a review of sensory substitution for the blind, Loomis and Klatzky [18] pointed to the need to couple device design to the capabilities of the user *for a given task*. Low bandwidth, for example, need not be a negative feature of a device if a task can competently be performed on the basis of the information provided. High bandwidth communication to the user is not an advantage if the target sensory channel cannot process the information provided. Furthermore, it should not be assumed that information-processing capacities of blind and sighted are equivalent; for example, superior tactile sensory systems may be preserved in older braille readers relative to sighted populations of the same age [19], whereas tasks that draw on visual imagery may be impeded in a blind population without any experience of sight.

## II. METHODS

### A. APPARATUS AND STIMULI

We have developed a series of working prototypes, leading up to the one used in the experiments described in the present paper. As the progression is informative, we review them briefly here. Our initial system permitted active interrogation of the visual surroundings with a miniature red laser attached to the fingertip (See Fig. 1) [20]. The laser was modulated at 10 KHz, allowing its reflection by objects in the environment to be detected by a non-directional phototransistor in the midst of other sources of light. Thus, like the Burch device, it sensed the light (in this case reflected) from a single point in space, but at a distance. Moreover, it differed from that device by actively interrogating across the detected point for an edge.

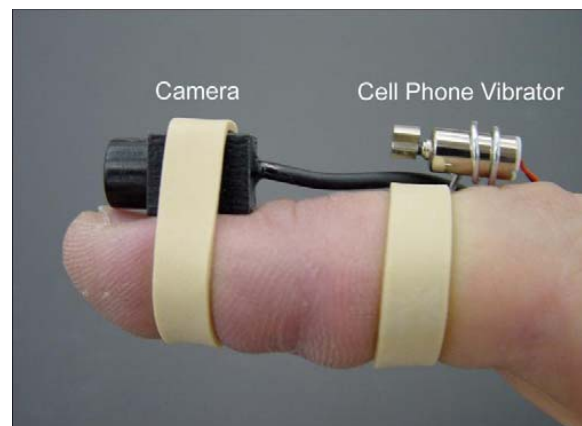


**FIGURE 1.** Initial laser-based FingerSight device for locating edges.

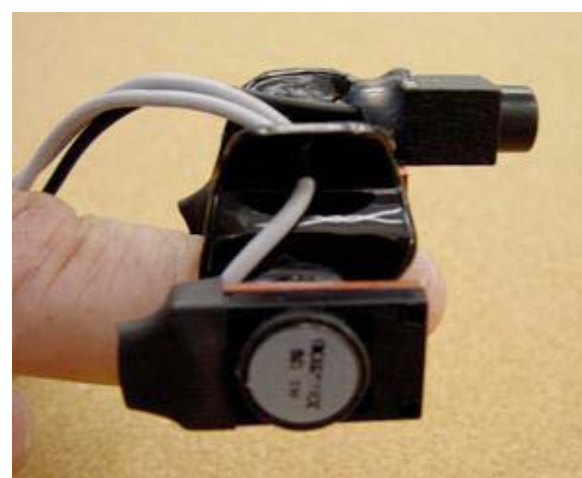
This was accomplished by means of a solenoid constructed from tiny magnets and a coil. Regenerative feedback between the amplitude of the detected signal from the reflected laser and the solenoid caused the laser to vibrate vertically when the laser spot was located on any edge between light and dark objects. Thus the system created a haptic stimulus (vibration) whenever a properly oriented visual edge was encountered by the laser.

The laser-based system was limited to simple edges and had the disadvantage that the active visible light source could be disturbing to others. More importantly, limiting FingerSight to what amounts to a single scanning line makes it extremely difficult to integrate more than the most rudimentary features in the visual field. A full video image captured at each pose of the finger offers many advantages, including immediate extraction of more complex features, identification of entire objects, and even determination of camera motion directly from the changing image. The recent availability of very inexpensive and small cameras, brought about largely by their use in cell phones, has led us to adopt a camera-based approach, the first version of which is shown in Fig. 2 [21]. A small black-and white camera (SuperCircuits PC206XP 510 × 492) was mounted on the subject’s finger along with a cell-phone vibrator. Real-time analysis of the video signal from a miniature camera was used to control the cell-phone vibrator. The cell-phone vibrator has the advantages of low cost and size, but presents problems in that amplitude and frequency are inextricably linked together by the operating voltage. Also, the frequency is low enough to cause noticeable vibration in the camera image, and the time for the motor to come up to speed is relatively long.

The next iteration of the FingerSight system is shown in Fig. 3 [22]. Two small speakers (GUI, Inc. #GC0251K-ND, 1W, 8Ω) were converted into haptic vibratory stimulators, or *tactors*, by cementing a short wooden dowel onto the central cones of each speaker. The dowel passed through a rubber grommet pressed against the subject’s skin. The converted



**FIGURE 2.** Simple camera-based FingerSight device with cell-phone vibrator.



**FIGURE 3.** FingerSight device with two speaker-based vibrators.

speakers were mounted on either side of the finger and held there, along with the camera, by a springed clip, leaving the palmar aspect of the fingertip bare to use for touching and grasping objects. We hoped, by including two such factors, to be able to use the relative strength of the vibration from each factor to convey a parameter such as location of an image feature along the horizontal axis.

This model did not perform well for identifying the location of visual targets, primarily due to mechanical design issues. It proved problematic to mount the device securely on the distal phalanx of the finger. Furthermore, given the influence of flexion/extension at the distal inter-phalangeal joint on the camera orientation, it was difficult for users to judge where they were aiming the device. In addition, there was low sensitivity to vibration the speakers mounted on the lateral aspect of the finger, as compared to the speaker on the medial aspect. Collectively, these issues made any asymmetry in signal strength from the two factors of little value to the user’s perception.

The current version of FingerSight uses the same speaker-based factor shown in Fig. 3, but with only one factor on

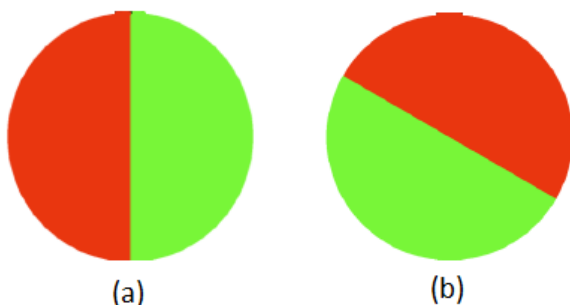
the (more sensitive) medial aspect of the finger, along with a miniature color camera (Supercircuits PC208,  $510 \times 492$ ). As shown in Fig. 4, these components were fitted into a 5-cm long aluminum splint. The camera extended approximately 8 mm from the tip of the finger, and the speaker was attached to the side of the splint. The splint weighed 18.8 g and caused minimal interference with movement of the finger as a whole, though it did restrict flexion between the phalanges, thereby reducing uncertainty in camera orientation. Covering the palmer aspect of the finger made it difficult to use the finger for other purposes, but we were willing to accept this for the present experiment. The width of the splint was adjustable to accommodate different finger sizes. The camera and the speaker were connected to a computer system, which performed spatiotemporal edge detection (described below) and controlled vibrator output. The haptic feedback was provided by a low-frequency (20 Hz) audio signal sent to the speaker, the amplitude of which was adjusted to produce detectable vibrations.



**FIGURE 4.** Current FingerSight device apparatus with camera and stimulator attached.

## B. IMAGE ANALYSIS

For the purposes of experimental testing, we constructed a circular target using simple computer graphics. The area of the circle was divided across its center, one semicircle containing red and one containing green pixels (see Fig. 5). The line across the center of the circle separating red from green pixels could be set to any desired orientation. The circular



**FIGURE 5.** Examples of edge stimuli, at 12 (a) and 4 (b) o'clock, respectively.

area was surrounded by white pixels and the images were projected on a screen, where the user could point the finger-mounted camera at it.

Software was developed to analyze the video image from the finger-mounted camera in real time and to control the vibratory feedback to that finger. The goal was to generate vibrations whenever the user's finger was pointing at, or had just passed over, the boundary between the red and green areas, irrespective of the edge's orientation. Optical edge detection was performed with a novel spatiotemporal kernel, designed to detect edges both in the spatial domain (when the camera was resting on an edge) and the temporal domain (when the camera had moved completely past an edge between one video frame and the next). This design accommodated fast hand movements, which might cause the edge to never actually have been within the kernel in any given frame, as well as slower continual tracking along the edge. The spatiotemporal kernel accomplished both of these for edges of any orientation by comparing each pixel in one time frame with the corresponding pixel directly across the center of a circular kernel in the next time frame. If the two pixels differed, an edge was present or had just been crossed.

It is important to note that unlike the general usage of the term "kernel" in image processing, especially for convolution, our spatiotemporal kernel was applied not at every pixel, but only at one location, the center of the image. Just as the fovea of the eye (the high resolution central region) is constantly moving for human vision to function, our kernel was physically moved by the user's finger along with the entire image, to interrogate the visual environment. The optimal size of the circular kernel varied depending on the scale of the edge features to be considered. For the particular experiment described here, the optimal radius of the kernel was determined experimentally to be 50 pixels.

A detailed formulation of the algorithm follows: We define  $K$  as the set of pixels within the circular kernel at the center of the image. Based on the particular red-green-blue (RGB) video image coming from the camera in real time, a color label  $c(x, y)$  was created for each pixel in  $K$ , denoting whether the pixel at location  $(x, y)$  was predominantly red, green, or ambiguous.

$$c(x, y) = \begin{cases} +1/2, & \text{if predominantly red} \\ -1/2, & \text{if predominantly green} \\ 0, & \text{if ambiguous} \end{cases} \quad (1)$$

For our purposes, "predominantly" indicates that the red (R) or blue (B) color value for the pixel in question was above a specific threshold, while the other two color values in the RGB pixel were both below their corresponding thresholds. Appropriate thresholds for red, green, and blue were determined experimentally, so that red, green, and white pixels (which registered as ambiguous) responded appropriately. We also created a binary mask value  $w(x, y)$  for each pixel that had the value of 1 wherever  $c$  was non-zero, so that boundaries including the white area surrounding the circle



could be excluded from generating edges.

$$w(x, y) = \begin{cases} 1 & \text{if } c(x, y) \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

For each pixel in  $K$ , we also considered another pixel in  $K$  directly across the origin (center of the kernel) at location  $(-x, -y)$ . Our measure of boundary strength was calculated by comparing the value  $c_t$  for the pixel at  $(x, y)$  in the current image at time  $t$  to value of  $c_{t-1}$  for the pixel at  $(-x, -y)$  in the previous image at time  $t - 1$ . The mask values  $w_t$  and  $w_{t-1}$  were used to ensure that no comparisons were made using “ambiguous” pixels. Thus a measure of edge strength  $d$  was computed as

$$d = \sum_{(x,y) \in K} |c_t(x, y) - c_{t-1}(-x, -y)| w_t(x, y) w_{t-1}(-x, -y) \quad (3)$$

If the edge measure  $d$  was above some threshold (determined experimentally), the software reported that an edge had been detected, and triggered the haptic feedback in the form of vibration at the fingertip.

The algorithm has the advantage that it measures equally well the conditions of sitting on an edge (of any orientation) for two successive timeframes and having passed completely over the edge from one timeframe to the next. The algorithm is also expandable to more than two dimensions. For example, in 3D one would simply compare pixels directly across the center of a *spherical* rather than a circular kernel.

### C. EXPERIMENTAL DESIGN

Twelve sighted subjects (10 males and 2 female) were recruited from the university population, with an average age of  $29.4 \pm 9.8$  years. All gave informed consent.

Subjects took part in a series of trials, in which they identified the angle of an edge stimulus according to a closed set of alternatives. Each edge stimulus, as described above, was projected on a display screen and consisted of a circular area 115 cm in diameter on a white background bisected with one half colored red and the other green. The angle of the bisection line could be varied (see Fig. 5). The circle was surrounded by white, so that the perimeter of the circle would not itself be interpreted as an edge (as described above).

Subjects were fitted with earplugs, blindfolded, and seated with the right arm resting on a foam support 79 cm above the floor. The FingerSight device was placed on the subject’s right index finger. The circles were projected (Epson PowerLite 70c) on a  $154 \times 115$  cm screen positioned 168 cm in front of the subject’s fingertip. The arm support was arranged so that the subject’s arm and finger were initially pointing perpendicular to the center of the projected image. During each trial, the subject freely explored the display edge with the FingerSight device as long as desired. Generally, subjects maintained their arm on the rest and moved their forearm and wrist. After exploration, the subject selected a response angle from the set of choices.

Two different sets of angles were presented, using different angular resolutions. Angle Set 1 consisted of six angles equally spaced over a  $180^\circ$  range in  $30^\circ$  steps (from positive vertical,  $90^\circ$ , to just short of negative vertical,  $-60^\circ$ ), corresponding to clock-hours from 12 to 5 o’clock. Angle Set 2 consisted of seven angles equally spaced over a  $90^\circ$  range in  $15^\circ$  steps (from positive vertical,  $90^\circ$ , to horizontal,  $0^\circ$ ) corresponding to the hours and half hours between 12 to 3 o’clock. Subjects reported their responses using these “clock face” labels. Within each set, the order of angles was randomized for each subject.

### III. RESULTS

Tables 1 and 2 show the confusion matrices for all subjects for Angle Sets 1 and 2, respectively, plotting reported angle vs. stimulus angle. A confusion matrix shows correct measurements along the diagonal, with increasing error as one moves off the diagonal. Overall, subjects had an accuracy rate (proportion correct) of  $0.77 \pm 0.05$  on Angle Set 1 and  $0.52 \pm 0.09$  on Angle Set 2 (95% confidence intervals). These levels are well above chance in this six-alternative choice task (chance = 0.16).

The data were also analyzed by measuring information transfer  $I$ , in order to determine the number of “bits” that can be transmitted by the device under the experimental conditions [23].

$$I = \sum_{j=1}^k \sum_{i=1}^k P(S_i, R_j) \log_2 \left( \frac{P(S_i|R_j)}{P(S_i)} \right). \quad (4)$$

TABLE 1. Confusion matrix for set 1.

		Reported Angle (deg)					
		90	60	30	0	-30	-60
Stimulus Angle (deg)	90	71	13	1	0	1	10
	60	1	68	25	2	0	0
	30	0	8	84	3	1	0
	0	1	0	10	81	3	1
	-30	0	0	0	6	81	9
	-60	4	0	0	0	32	60

Confusion matrix for Angle Set 1. Each entry is the number of occurrences of the associated stimulus/response pair.

TABLE 2. Confusion matrix for set 2.

		Reported Angle (deg)						
		0	15	30	45	60	75	90
Stimulus Angle (deg)	0	64	25	3	2	2	0	0
	15	19	49	24	2	1	0	1
	30	1	20	55	19	1	0	0
	45	0	8	36	39	12	1	0
	60	0	1	15	23	46	9	2
	75	0	0	1	15	33	38	9
	90	0	0	1	1	11	25	58

Confusion matrix for Angle Set 2. Each entry is the number of occurrences of the associated stimulus/response pair.

Here,  $k$  is the number of possible stimuli, and  $S_i$  and  $R_j$  represent a specific stimulus-response pair. This statistic can be estimated from the confusion matrices according to Equation 6.

$$I_{est} = \sum_{j=1}^k \sum_{i=1}^k \left( \frac{n_{ij}}{n} \right) \log_2 \left( \frac{n_{ij} \times n}{n_i \times n_j} \right). \quad (5)$$

Here,  $n_{ij}$  is the number of joint occurrences of stimulus  $i$  and response  $j$ ,  $n_i$  is the overall occurrence of stimulus  $i$ ,  $n_j$  is the overall occurrence of stimulus  $j$ , and  $n$  is the total number of trials.

Relative to the 2.58 bits of information in the six-alternative choice task, Angle Set 1 (angles with  $30^\circ$  increments) had an information transfer of 1.62 bits, while Angle Set 2 (angles with  $15^\circ$  increments) had an information transfer of 1.12 bits. These constitute transmission of 63% and 43% of the available information, respectively. This measure does not reflect the fact that most confusions were between adjacent angles.

#### IV. CONCLUSION

We have developed a new method of providing the finger with haptic feedback about visual targets in the general environment, by permitting the finger to scan the environment directly to locate those targets. We have developed this idea, which we call FingerSight, through a series of prototypes, exploring the particular application of detecting edges. In the process, we have developed a new spatiotemporal  $n$ -dimensional edge detection algorithm of potentially general interest, which is simultaneously capable of detecting the presence of a stationary edge as well as having passed over an edge between successive timeframes.

We have conducted proof-of-concept experiments showing that under active exploration the FingerSight device is capable of transmitting visual spatial information through haptic channels. Sighted, blindfolded subjects were able to use the device to differentiate between angles separated by as little as  $15^\circ$ . Performance with the coarser angle set, where responses were separated by  $30^\circ$  showed that FingerSight transmitted close to  $2/3$  of the available information in the stimulus, with most errors being near misses.

The information transmission of the FingerSight device is intrinsically constrained by basic sensorimotor abilities of motor control, kinesthetic sensing, and spatial cognitive processing. The latter two processes may have substantial impact on performance. In [9] it was found that when blindfolded, sighted adults tried to report the orientation of a raised line that was easily tracked by touch on the plane of a tabletop, their responses pulled the true value towards the sagittal axis by about 25% (plus some constant error). Since motor control was minimized in the task in [9], errors can be entirely attributed to the processes of kinesthetic sensing and building a spatial representation. The observed level of distortion is by itself substantial enough to lead to confusion errors in the present task.

These observations suggest that performance with FingerSight could be improved by features that support motor control and augment kinesthetic feedback. The present experiments measure only the performance of blindfolded sighted subjects who are novices using the device; training would be expected to considerably improve performance. Siegle and Warren have shown that distal attribution, i.e. the direct perception of objects in external space, can occur with similar finger-mounted sensory substitution systems after a period of training [24]. However the device used in their experiments had the vibrotactile stimulator mounted on the subject's back. It is possible that training time may be decreased with a device such as FingerSight, where the stimulus is more strongly tied to the subject's proprioceptive knowledge of the hand's location in space.

It is interesting to compare the accuracy of angular measurements by FingerSight in 3D space with other researchers who restricted interrogation to a fixed plane. Kennedy, et al., explored the ability to recognize 90- or 180-degree rotation on raised-outline drawings directly touched by the blind or sighted-blindfolded subjects [25], but did not explore finer angular resolution. Postma, et al., demonstrated the role of visual experience in such haptic spatial tasks, including the description of angles between bars on a table, showing that blindfolded *sighted* subjects outperformed *late blind* subjects, who outperformed *early blind* subjects [26]. They found that having experienced vision, even to a limited extent, helps in the interpretation of angle by touch alone. The worst performance in verbally judging the angle of the bars (demonstrated by the early blind) was 7.2 degrees, a good deal better than our errors, possibly because they were constrained to a tabletop. It would also be interesting to see if sighted individuals also have a similar advantage in learning to use FingerSight over blind individuals, due to previous visual experience.

Also relevant to the question of angle and touch is the work of Rastogi et al., with haptic computer mice, which are modified to have tactile pin arrays on their upper surface [27]. They report a "significant lack of accuracy in the haptic position information, which is critical for individuals to haptically piece together a 2-D graphic." The inaccuracy is due to the fact that the tactile mouse (or any normal mouse) is a relative positioning device, dependent upon the speed of motion and orientation of the mouse to determine total displacement on the screen. In contrast, FingerSight is inherently an absolute positioning device, given a stationary environment, and as such, FingerSight may have an advantage.

The system of Burch and Pawluk previously mentioned [17] uses a fingertip photosensor and piezoelectric stimulator to scan specially created graphical displays, in which texture is added to enhance perception of edges and orientations. A single photosensor thus suffices for this purpose, only because preprocessing is performed to populate regions on either side of boundaries with differing textures. Multiple photosensors on different fingers were found to improve results with this system [28], because the operator's

knowledge of the spatial relationship between the fingertips could be used to integrate the inputs. But still the approach relies on preprocessing the image to create textures, after which individual photosensors can be effective. With FingerSight, we are exploring the unadulterated 3D environment, where depending on only a few photosensors is not sufficient. We benefit greatly by having a multi-pixel image at each timeframe.

By further processing of the camera images, the FingerSight device could be adapted to identify more complicated features than simple visual edges, perhaps even constituting an object recognition system for a blind operator. For example, it could be used to find a particular person in a crowd, or identify a door displaying a particular sign. The rapid advancement in computer vision algorithms, driven in part by the security and social networking industries, will provide ever more sophisticated capabilities. For example, determining camera motion and target depth from a image sequence could permit greater 3D integration of multiple perspectives and facilitate providing navigational cues to the blind operator, such as where the curbside is, whether the approaching stairwell goes up or down, or whether one is moving towards or away from the elevator. Such analyses are not feasible for a single photosensor, but rather, they require an entire image.

A common problem in any real-world computer vision application is the variability of lighting, and there are established techniques for solving this problem using such constructs as the “Illumination Cone” [29]. Another approach is to use an infrared camera with its own lighting source. Some of these go beyond simple 2D image formation. For example, a recently developed Time-of-Flight (TOF) 3D camera, the Swiss Ranger SR4000 (MESA Imaging, Zuerich), can deliver a  $176 \times 144$  pixel image with each pixel reporting range up to 10 m with 1 cm accuracy. The present camera is roughly  $7 \text{ cm}^3$  and is being integrated by at least one research group into portable devices for the blind [30]. When further miniaturized, it may provide true 3D data for FingerSight.

For some of these more sophisticated systems, it may become problematic to rely solely on vibrotactile feedback to the finger. The bandwidth of audio, especially combined with language, makes it an appealing option, though as noted above, one does not want to impede the natural use of auditory cues that are especially crucial for the vision impaired. However, intermittent use of verbal output by FingerSight, and for that matter, verbal commands by the operator to the device, could prove extremely useful, while not impeding auditory cues from the environment.

As noted above, it is imperative to test whether features added to the technology ultimately add to the functional capability of the user, given intrinsic limitations on human information processing. That said, one promising avenue is the incorporation of control capabilities into the FingerSight device. One of our previous systems included the capability to

control graphical objects on a screen. In this implementation, the location of a small white square on a screen is controlled by motion of the finger. The algorithm detects the square in the field of view of the camera, moving it on the screen to keep it constantly in the center of the camera image, while providing haptic feedback about whether the tracking system has locked onto the target. A variation on this system constrains the square to move along a straight line, simulating the action of a slide pot under the operator’s control. A further variation uses a small white triangle to simulate a knob, whose orientation is determined using standard computer vision techniques and subsequently controlled by rotation of the finger [22]. Clearly, such systems are not limited to actively controlling graphical objects on a screen, but could also identify inanimate objects such as a light switch or door latch. In such cases, remote control could still be achieved by motion of the finger once the target has been identified, using a separate control channel to turn on the light or lock the door. Continued work on the control aspects of FingerSight is an integral part of our plan for future development of the device.

A final note is merited on the eventual miniaturization of a FingerSight device such that it might be small enough to be worn like an artificial fingernail. The cameras themselves are almost small enough already, and the main considerations are power and communications. One can envision radio communication between a fingertip device and a pocket unit, much as wireless earphones and microphones communicate with cell phones now. Such devices might be fully integrated into the everyday activities of the vision-impaired.

## REFERENCES

- [1] E. R. Strelow and D. H. Warren, *Electronic Spatial Sensing for the Blind*. Dordrecht, The Netherlands: Martinus Nijhoff Publishers, 1985.
- [2] J. L. Loomis R. G. Golledge, R. L. Klatzky, and J. R. Marston, “Assisting wayfinding in visually impaired travelers,” in *Applied Spatial Cognition: From Research to Cognitive Technology*, G. Allen Ed. Mahwah, NJ, USA: Lawrence Erlbaum Associates, 2007, pp. 179–202.
- [3] B. Bentzen, W. Crandall, and L. Myers, “Use of infrared remote signage in a complex transit station by persons with developmental delays or dyslexia,” *J. Vis. Impairment, Blindness*, vol. 91, no. 1, pp. 407–10, 1997.
- [4] J. Sibert and M. Gokturk, “Direct pointing apparatus and method therefor,” U.S. Patent 6 184 863, Feb. 6, 2001.
- [5] R. G. Golledge, J. Marston, J. M. Loomis, and R. L. Klatzky, “Stated preferences for components of a personal guidance system for nonvisual navigation,” *J. Vis. Impairment Blindness*, vol. 98, no. 3, pp. 135–147, 2004.
- [6] C. Lenay, O. Gapenne, S. Hanne-ton, C. Marque, and C. Genouelle, “Sensory substitution: Limits and perspectives,” in *Touching For Knowing: Cognitive Psychology Of Haptic Manual Perception*, Y. Hatwell, A. Steri, and E. Gentaz, Eds. Amsterdam, The Netherlands: John Benjamins, 2003, pp. 275–292.
- [7] J. M. Loomis, “Sensory replacement and sensory substitution: Overview and prospects for the future,” in *Converging Technologies for Improving Human Performance: Nanotechnology, Biotechnology, Information Technology and Cognitive Science*, M. C. Roco and W. S. Bainbridge Eds. Dordrecht, The Netherlands: Kluwer, 2003, pp. 213–223.
- [8] J. J. Gibson, *The Senses Considered as Perceptual Systems*. Boston, MA, USA: Houghton Mifflin, 1996.

- [9] R. Velázquez, E. E. Pissaloux, and M. Wiertelowski, "A compact tactile display for the blind with shape memory alloys," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2006, pp. 3905–3910.
- [10] F. Maingreud, E. E. Pissaloux, R. Velazquez, F. Gaunet, M. Hafez, and J. M. Alexandre, "A dynamic tactile map as a tool for space organization perception: Application to the design of an electronic travel aid for visually impaired and blind people," in *Proc. IEEE Eng. Med. Biol. Soc.*, Jan. 2005, pp. 6912–6915.
- [11] P. B.-Y. Rita and M. E. Tyler, "Tongue man-machine interface," in *Proc. Med. Meets Virtual Reality*, 2000, pp. 17–19.
- [12] K. Nunokawa and S. Ino, "An experimental study on target recognition using white canes," in *Proc. IEEE Eng. Med. Biol. Soc.*, Aug. 2010, pp. 6583–6586.
- [13] P. B. L. Meijer, "An experimental system for auditory image representations," *IEEE Trans. Biomed. Eng.*, vol. 39, no. 2, pp. 112–121, Feb. 1992.
- [14] S. Lederman and R. Klatzky, "Haptic perception: A tutorial," *Attention, Perception, Psychophys.*, vol. 71, no. 7, pp. 1439–1459, 2009.
- [15] J. C. Craig, "Vibrotactile pattern perception: Extraordinary observers," *Science*, vol. 196, pp. 450–453, Apr. 1977.
- [16] C. Lenay, S. Canu, and P. Villon, "Technology and perception: The contribution of sensory substitution systems," in *Proc. Int. Conf. Cognit. Technol.*, 1997, pp. 44–53.
- [17] D. Burch and D. Pawluk, "A cheap, portable haptic device for a method to relay 2-D texture-enriched graphical information to individuals who are visually impaired," in *Proc. 11th Int. ACM SIGACCESS Conf. Comput. Accessibility*, Pittsburgh, PA, USA, 2009, pp. 215–217.
- [18] J. L. Loomis, R. L. Klatzky, and N. A. Giudice, "Sensory substitution of vision: Importance of perceptual and cognitive processing," in *Assistive Technology for Blindness and Low Vision*, R. Manduchi and S. Kurniawan Eds. Boca Raton, FL, USA: CRC Press, 2012, pp. 162–191.
- [19] G. E. Legge, C. Madison, B. N. Vaughn, A. M. Cheong, and J. C. Miller, "Retention of high tactile acuity throughout the life span in blindness," *Perception, Psychophys.*, vol. 70, no. 8, pp. 1471–1488, 2008.
- [20] K. Zawrotny, A. Craig, D. Weiser, R. L. Klatzky, and G. D. Stetten, "Fingertip vibratory transducer for detecting optical edges using regenerative feedback," in *Proc. 14th Symp. Haptic Int. Virtual Environ. Teleoperator Syst.*, Arlington, TX, USA, Mar. 2006, pp. 373–374.
- [21] G. Stetten et al., "Fingersight: Fingertip visual haptic sensing and control," in *Proc. IEEE Int. Workshop Haptic Audio Vis. Environ. Appl.*, Oct. 2007, pp. 80–83.
- [22] J. Galeotti, S. Horvath, R. Klatzky, B. Nichol, M. Siegel, and G. Stetten, "FingerSight: Fingertip control and haptic sensing of the visual environment," *SIGGRAPH New Tech. Demos*, Los Angeles, CA, USA, Aug. 2008, pp. 11–15.
- [23] H. Tan, H. Z. Tan, N. I. Durlach, W. M. Rabinowitz, and C. M. Reed, "Information transmission with a multifinger tactual display," *Perception, Psychophys.*, vol. 61, no. 6, pp. 993–1008, 1999.
- [24] J. H. Siegle and W. H. Warren, "Distal attribution and distance perception in sensory substitution," *Perception*, vol. 39, no. 2, pp. 208–223, 2010.
- [25] J. M. Kennedy and J. Bai, "Haptic pictures: Fit judgments predict identification, recognition memory, and confidence," *Perception*, vol. 31, no. 8, pp. 1013–1026, 2002.
- [26] A. Postma, S. Zuidhoek, M. L. Noordzij, and A. M. Kappers, "Haptic orientation perception benefits from visual experience: Evidence from early-blind, late-blind, and sighted people," *Perception, Psychophys.*, vol. 70, pp. 1197–1206, 2008.
- [27] R. Rastogi, D. Pawluk, and J. Ketchum, "Issues of using tactile mice by individuals who are blind and visually impaired," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 3, pp. 311–318, Jun. 2010.
- [28] D. S. Burch and D. T. V. Pawluk, "Using multiple contacts with texture-enhanced graphics," in *Proc. World Haptics*, Istanbul, Turkey, Jun. 2011, pp. 21–24.
- [29] P. N. Belhumeur and D. J. Kriegman, "What is the set of images of an object under all possible lighting conditions?" in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1996, pp. 270–277.
- [30] C. Ye, "Navigating a portable robotic device by a 3D imaging sensor," in *Proc. Sensors IEEE*, Nov. 2010, pp. 1005–1010.



**SAMANTHA HORVATH** received the M.S. degree in robotics from Carnegie Mellon University, where she is currently pursuing the Doctoral degree with the Robotics Institute. Her research interests include medical device design and medical image analysis, with a focus on computer vision and optical engineering. She has done research on image-guided ultrasound and augmented reality systems for surgical microscopes.



**JOHN GALEOTTI** received the Ph.D. degree in robotics from Carnegie Mellon University. He is a Senior Project Scientist in robotics and an Adjunct Assistant Professor in biomedical engineering with Carnegie Mellon University, and an Adjunct Assistant Professor in bioengineering with the University of Pittsburgh. His research interests include both biomedical optics and biomedical image analysis and visualization, in particular, for intraoperative guidance. He has done research applying real-time computer-controlled optics, image analysis, and visualization approaches to applications ranging from computer-vision for tracking ultrasound probes in anatomical coordinates to real-time microsurgical guidance systems utilizing optical coherence tomography.



**BING WU** received the Ph.D. degree in experimental psychology from the University of Louisville. He is an Assistant Professor with the College of Technology and Innovation, Arizona State University. He worked as a Post-Doctoral Researcher with Drs. Klatzky and Stetten at Carnegie Mellon University and the University of Pittsburgh, from 2004 to 2011. His research interests include spatial perception/cognition, the visual and haptic control of action, and human factors issues in real, teleoperated, or simulated medical operations.



**ROBERTA KLATZKY** received the Ph.D. degree in psychology from Stanford University. She is a Professor of psychology and human-computer interaction with Carnegie Mellon University. Her research interests include human perception and cognition, with special emphasis on spatial cognition and haptic perception. She has done extensive research on human haptic and visual object recognition, navigation under visual and nonvisual guidance, and perceptually guided action, with application to navigation aids for the blind, haptic interfaces, exploratory robotics, image-guided surgery, and virtual environments.





**MEL SIEGEL** received the Ph.D. degree in physics from the University of Colorado Boulder. He is a Faculty Member in robotics and an Affiliated Faculty Member in human–computer interaction with Carnegie Mellon University. His research interests include sensing, sensors, perception, and display systems in robotics contexts. He has done extensive research in sensors for robot proprioception and environmental awareness, robots for sensing missions, 3-D stereoscopic display systems, and scaling, power, and energy issues in robotics. In addition to his research and teaching, he also directs the Master of Science in Robotics Technology program. He has been an Active Member and Officer of the Instrumentation and Measurement Society.



**GEORGE STETTEN** received the M.D. and Ph.D. degrees in biomedical engineering from SUNY Syracuse and UNC Chapel Hill, respectively. He is a Professor of bioengineering with the University of Pittsburgh and Research Professor with the CMU Robotics Institute. He is the Director with the Visualization and Image Analysis Laboratory and the Music Engineering Laboratory. His research interests include image guided intervention, haptics, and image analysis. He was a founding contributor to the National Library of Medicine Insight Toolkit and is a fellow of the American Institute for Medical and Biological Engineering.