# The Design and Preliminary Evaluation of a Finger-Mounted Camera and Feedback System to Enable Reading of Printed Text for the Blind

**Lee Stearns[1], Ruofei Du[1], Uran Oh[1], Yumeng Wang[2], Leah Findlater[3], Rama Chellappa[4], Jon E. Froehlich[1]**

[1]Computer Science, [2]School of Architecture,
[3]College of Information Studies, [4]Electrical and Computer Engineering
University of Maryland, College Park
{lstearns, ruofei, uranoh, yumeng, leahkf, chella, jonf}@umd.edu

**Abstract.** We introduce the preliminary design of a novel vision-augmented touch system called *HandSight* intended to support activities of daily living (ADLs) by sensing and feeding back non-tactile information about the physical world *as it is touched*. Though we are interested in supporting a range of ADL applications, here we focus specifically on reading printed text. We discuss our vision for HandSight, describe its current implementation and results from an initial performance analysis of finger-based text scanning. We then present a user study with four visually impaired participants (three blind) exploring how to continuously guide a user's finger across text using three feedback conditions (haptic, audio, and both). Though preliminary, our results show that participants valued the ability to access printed material, and that, in contrast to previous findings, audio finger guidance may result in the best reading performance.

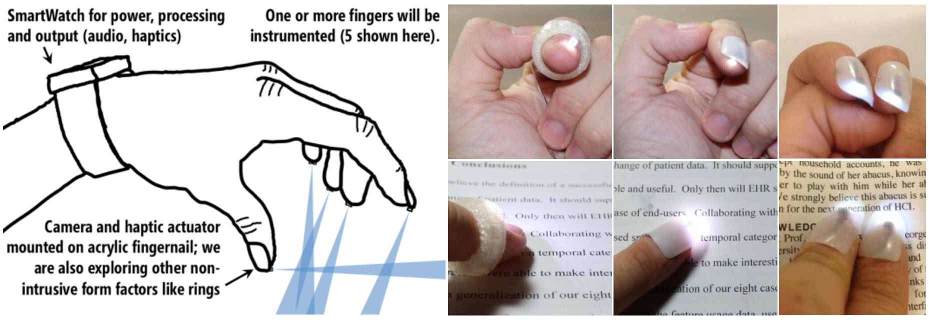
**Keywords:** Accessibility, Wearables, Real-time OCR, Text Reading for Blind

## 1 Introduction

Over 285 million people have visual impairments (VI) worldwide—including 39 million who are blind—that can affect their ability to perform activities of daily living [1]. In many countries, such as the US [2], VI prevalence is increasing due to aging populations. While previous research has explored combining mobile cameras and computer vision to support people with VI for at-a-distance information tasks such as navigation (*e.g.*, [3]–[8]), facial recognition (*e.g.*, [9]–[12]), and spatial perception (*e.g.*, [13]–[15]), they fail to support proximal information accessed through touch.

We are pursuing a new approach: a vision-augmented touch system called *HandSight* that supports activities of daily living (ADLs) by sensing and feeding back non-tactile information about the physical world *as it is touched*. Although still at an early stage, our envisioned system will consist of tiny CMOS cameras (1×1mm$^2$) and micro-haptic actuators mounted on one or more fingers, computer vision and machine learn-

adfa, p. 1, 2011.


SmartWatch for power, processing and output (audio, haptics)
One or more fingers will be instrumented (5 shown here).
Camera and haptic actuator mounted on acrylic fingernail; we are also exploring other non-intrusive form factors like rings

**(a)** Example HandSight Envisionment (w/5 instrumented fingers) **(b)** Ring Form Factor **(c)** Nail Form Factor **(d)** 2-Finger Setup

**Figure 1: HandSight uses a 1×1mm$^2$ *AWAIBA NanEye 2C* camera developed for minimally invasive surgeries (*e.g.*, endoscopies) that can capture 250×250px images at 44fps. The above images: (a) a design mockup and (b-d) early form factors with the NanEye camera. In this paper, we explore a single camera implementation with a ring form factor in a text reading context (see Figure 2).**

ing algorithms to support fingertip-based sensing, and a smartwatch for processing, power, and speech output; see Figure 1. Since touch is a primary and highly attuned means of acquiring information for people with VI [16], [17], we hypothesize that collocating the camera with the touch itself will enable new and intuitive assistive applications.

While we are interested in supporting a range of applications from object recognition to color identification, in this paper we focus on the challenge of reading printed text. Our overarching goal is to allow blind users to touch printed text and receive speech output in real-time. The user's finger is guided along each line via haptic and non-verbal audio cues. At this stage, our research questions are largely exploratory, spanning both human-computer interaction (HCI) and computer vision: How can we effectively guide the user's finger through haptic and auditory feedback to appropriately scan the target text and notify them of certain events (*e.g.*, start/end of line or paragraph reached)? How accurately can optical character recognition (OCR) be achieved at a speed that is responsive to the user's touch? How do the position, angle, and lighting of the finger-mounted camera affect OCR performance?

To begin examining these questions, we pursued two parallel approaches. For the computer vision questions, we developed an early HandSight prototype along with efficient algorithms for perspective and rotation correction, text detection and tracking, and OCR. We present preliminary evaluations and demonstrate the feasibility of our envisioned system. For the HCI-related questions, we developed a custom test apparatus on the Apple iPad that simulates the experience of using HandSight, but provides us with additional experimental control and allows us to more precisely track the user's finger in response to feedback conditions. Using this setup, we report on a preliminary evaluation with four VI participants (three blind) across three finger guidance conditions: *audio*, *haptics*, and *audio+haptics.* Findings suggest that audio may be the most intuitive feedback mechanism of the three.

Compared to the majority of emerging computer vision systems to support VI users, which use head- or chest-mounted cameras (*e.g.,* [4], [12], [18]), our system offers two primary advantages: (i) *collocation of touch, sensing, and feedback*, potentially enabling more intuitive interaction and taking advantage of a VI individual's high tactile acuity [16], [17]; (ii) *unobtrusive, always-available interaction* that allows for seamless switching between the physical world and vision-augmented applications. As a finger-mounted approach, our work is most similar to [19]–[21], described next.

## 2 Related Work

Scientists have long sought to support blind people in reading printed text (for a review, see [22], [23]). Many early so-called "reading machines for the blind" used a sensory substitution approach where the visual signals of words were converted to non-verbal auditory or tactile modalities, which were complicated to learn but accessible. Two such examples include the *Optophone*, which used musical chords or 'motifs' [24] and the *Optacon*, which used a vibro-tactile signal [25], [26]. With advances in sensing, computation, and OCR, modern approaches attempt to scan, recognize, and read aloud text in real-time. This transition to OCR and speech synthesis occurred first with specialized devices (*e.g.,* [27]–[29]), then mobile phones (*e.g.,* [30], [31]), and now wearables (*e.g.,* [12], [21]). While decades of OCR work exist (*e.g.,* [32]–[35]), even state-of-the-art reading systems become unusable in poor lighting, require careful camera framing [36], [37], and do not support complex documents and spatial data [38]. Because HandSight is self-illuminating and co-located with the user's touch, we expect that many of these problems can be mitigated or even eliminated.

As a wearable solution, HandSight is most related to *OrCam* [12] and *FingerReader* [21]. OrCam is a commercial head-mounted camera system designed to recognize objects and read printed text in real-time (currently in private beta testing). Text-to-speech is activated by a pointing gesture in the camera's field-of-view. While live demonstrations with sighted users have been impressive (*e.g.,* [39], [40]), there is no academic work examining its effectiveness with VI users for reading tasks. The primary distinctions between HandSight and OrCam are, first, hand-mounted versus head-mounted sensing, which could impact camera framing issues and overall user experience. Second, HandSight supports direct-touch scanning compared to OrCam's indirect approach, potentially allowing for increased control over what is read and reading speed as well as increased spatial understanding of a page/object. Regardless, the two approaches are complementary, and we plan to explore a hybrid in the future.

More closely related to HandSight, FingerReader [21] is a custom finger-mounted device with vibration motors designed to read printed text by direct line-by-line scanning with the finger. Reported evaluations [21] of FingerReader are limited to a very small OCR assessment under unspecified "optimal" conditions, and a qualitative user study with four blind participants. The participants preferred haptic to audio-based finger guidance; this finding is the opposite of our own preliminary results, perhaps due to differences in how the audio was implemented (theirs is not clearly described). Further, our study extends [21] in that we also present user performance results.

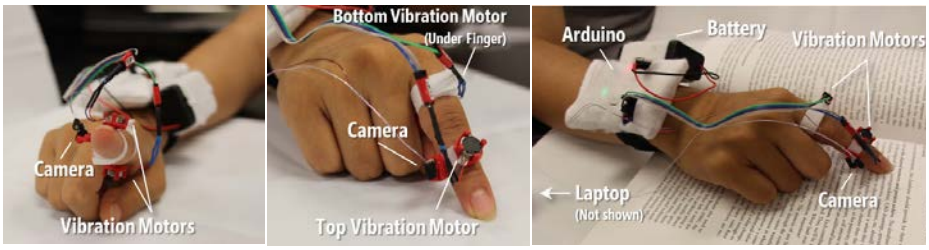
In terms of finger guidance, haptic and audio feedback have been used in numerous projects to guide VI users in exploring non-tactile information or tracing shapes. Crossan and Brewster [41], for example, combined pitch and stereo sonification with a force feedback controller to drag the user along a trajectory, and found that performance was higher with audio and haptic feedback than haptic feedback alone. Other approaches have included sonification and force feedback to teach handwriting to blind children [42], speech-based icons or "spearcons" [43], vowel sounds to convey radial direction [44], and use of primarily tactile feedback to transmit directional and shape data [45]–[47]. Our choice to vary pitch for audio-based line tracing feedback with HandSight is based on previous findings [41], [43], [48]. Oh *et al.* [48], *e.g.*, used sonification to support non-visual learning of touchscreen gestures; among a set of sound parameters tested (pitch, stereo, timbre, etc.), pitch was the most salient.

## 3 System Design

HandSight is comprised of three core components: sensors, feedback mechanisms, and a computing device for processing. Our current, early prototype is shown in Figure 2. Before describing each component in more detail, we enumerate six design goals.

### 3.1 Design Goals

We developed the following design goals based on prior work and our own experiences developing assistive technology: (1) *Touch-based rather than distal interaction.* Although future extensions to HandSight could examine distal interaction, our focus is on digitally augmenting the sense of touch. (2) *Should not hinder normal tactile function.* Fingers are complex tactile sensors [49], [50] that are particularly attuned in people with visual impairments [16], [17]; HandSight should not impede normal tactile sensation or hand function. (3) *Easy-to-learn/use.* Many sensory aids fail due to their complexity and high training requirements [22]; to ensure HandSight is approachable and easy to use, we employ an iterative, human-centered design approach. (4) *Always-available.* HandSight should allow for seamless transitions between its use and real-world tasks. There is limited prior work on so-called always-available input [20], [51]–[53] for blind or low-vision users. (5) *Comfortable & robust.* HandSight's physical design should support, not encumber, everyday activities. It should be easily removable, and water and impact resistant. (6) *Responsive & accurate.* HandSight



(a) Close-up front view (b) Close-up side view (c) Full system view

**Figure 2: The current HandSight prototype with a NanEye ring camera, two vibration motors, and an Arduino. Finger rings and mounts are constructed from custom 3D-printed designs and fabric. Processing is performed in real-time on a laptop (not shown).**

should allow the user to explore the target objects (*e.g.*, utility bills, books) quickly—the computer vision and OCR algorithms should work accurately and in real-time.

## 3.2 Hardware

**Sensing Hardware.** Our current prototype uses a single 1×1mm$^2$ *AWAIBA NanEye 2C* camera [54] that can capture 250×250 resolution images at 44 frames per second (fps). The NanEye was originally developed for minimally invasive surgical procedures such as endoscopies and laparoscopies and is thus robust, lightweight, and precise. The camera also has four LEDs coincident with the lens (2.5mm ring), which enables dynamic illumination control. The small size allows for a variety of finger-based form factors including small rings or acrylic nail attachments. In our current prototype, the camera is attached to an adjustable velcro ring via a custom 3D-printed clip.

**Processing.** For processing, we use a wrist-mounted Arduino Pro Micro with an attached Bluetooth module that controls the haptic feedback cues. The video feed from the camera is currently processed in real time on a laptop computer (our experiments used a Lenovo Thinkpad X201 with an Intel Core i5 processor running a single computation thread at approximately 30fps). Later versions will use a smartwatch (*e.g.*, Samsung Galaxy Gear [55]) for power and processing.

**Feedback.** HandSight provides continuous finger-guidance feedback via vibration motors, pitch-controlled audio, or both. Our current implementation includes two vibration motors that are **8mm diameter disks and 3.4mm thick** (Figure 2), though we are actively looking at other solutions (see Discussion). A text-to-speech system is used to read each word as the user's finger passes over it, and distinctive audio and/or haptic cues can be used to signal other events, such as end of line, start of line, etc.

## 3.3 CV Algorithm Design and Evaluation

Our current HandSight implementation involves a series of frame-level processing stages followed by multi-frame merging once the complete word has been observed.

Below, we describe our five stage OCR process and some preliminary experiments evaluating performance.

**Stage 1: Preprocessing. We acquire grayscale video frames at ~40fps and 250x250px resolution from the NanEye camera (Figure 3).** With each video frame, we apply four preprocessing algorithms: first, to correct radial and (slight) tangential distortion, we use standard camera calibration algorithms [56]. Second, to control lighting for the next frame, we optimize LED intensity using average pixel brightness and contrast. Third, to reduce noise, perform binarization necessary for OCR, and adapt to uneven lighting from the LED, we filter the frame using an adaptive threshold in a Gaussian window; finally, to reduce false positives, we perform a connected component analysis and remove components with areas too small or aspect ratios too narrow to be characters.

**Stage 2: Perspective and Rotation Correction.** The finger-based camera is seldom aligned perfectly with the printed text (*e.g.*, top-down, orthogonal to text). We have observed that even small amounts of perspective distortion and rotation can reduce the accuracy of our text detection and OCR algorithms. To correct perspective and rotation effects, we apply an efficient approach detailed in [56]–[58], which relies on the parallel line structure of text for rectification. We briefly describe this approach below.

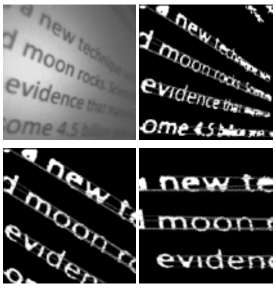


**Figure 3: A demonstration of our perspective and rotation correction algorithm.**

To identify potential text baselines, we apply a Canny filter that highlights character edges and a randomized Hough transform that fits lines to the remaining pixels. From this, we have a noisy set of candidate baselines. Unlikely candidates are filtered (*e.g.*, vertical lines, intersections that imply severe distortion). The remaining baselines are enumerated in pairs; each pair implies a potential rectification, which is tested on the other baselines. The baseline pair that results in the lowest line angle variance is selected and the resulting rectification is applied to the complete image.

More precisely, the intersection of each pair of baselines implies a horizontal vanishing point $V_x = l_1 \times l_2$ in homogeneous coordinates. If we assume the ideal vertical vanishing point $V_y = [0, 1, 0]^T$, then we can calculate the homography, $H$, that will make those lines parallel. Let $l_\infty = V_x \times V_y = [a, b, c]^T$ and calculate the perspective homography, $H_p$, using those values. The perspective homography makes the lines parallel, but does not align them with the *x*-axis. We must rotate the lines by an angle $\theta$ using a second matrix, $H_r$. The complete rectifying homography matrix becomes:

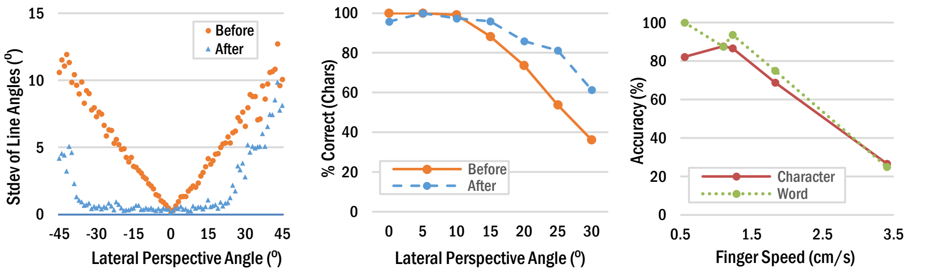
$$H = H_r H_p = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ a/c & b/c & 1 \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ a/c & b/c & 1 \end{bmatrix} \quad \mathbf{(1)}$$

To investigate the effect of lateral perspective angle on performance, we performed a synthetic experiment that varied the lateral angle from -45° to 45° across five randomly selected document image patches. The raw rectification performance is shown in Figure 4a and the effect of rectification on character-level OCR is shown in Figure 4b accuracy (the algorithm for OCR is described below).

Stage 3: Text Detection. The goal of the text detection stage is to build up a hierarchy of text lines, words, and characters. This task is simplified because we assume the perspective and rotation correction in Stage 2 has made the text parallel to the *x*-axis. First, we split the image into lines of text by counting the number of text pixels in each row and searching for large gaps. Next, we split each line into words using an identical process on the columns of pixels. Gaps larger than 25% of the line height are classified as spaces between words. Finally, we segment each word into individual characters by searching for local minima in the number of text pixels within each column.

**Stage 4: Character Classification.** Real-time performance is important for responsive feedback, which prevents us from using established OCR engines such as Tesseract. Thus, we compute efficient character features (from [59]), and perform classification using a support vector machine (SVM). Each character candidate is centered and scaled to fit within a 32x32 pixel window, preserving the aspect ratio. The window is split into four horizontal and vertical strips, which are summed along the short axis to generate eight vectors of length 32 each. These vectors, along with the aspect ratio, perimeter, area, and thinness ratio make up the complete feature vector. The thinness ratio is defined as $T=4\pi(A/P^2)$ where $A$ is the area and $P$ is the perimeter. We compensate for the classifier's relatively low accuracy by identifying the top $k$ most likely matches. By aggregating the results over multiple frames, we are able to boost performance.

**Stage 5: Tracking and final OCR result output.** The camera's limited field of view means that a complete word is seldom fully within a given frame. We must track the characters between frames and wait for the end of the word to become visible before we can confidently identify it. Character tracking uses sparse low-level features for efficiency. First, we extract FAST corners [60], and apply a KLT tracker [61] at their locations. We estimate the homography relating the matched corners using the random sample consensus [62]. After determining the motion between frames, we relate the lines, words, and individual characters by projecting their locations in the previous frame to the current frame using the computed homographies. The bounding boxes with the greatest amount of overlap after projection determine the matches. When the



**(a)** Performance of lateral perspective and rotation rectification algorithm.

**(b)** Effect of lateral perspective angle on accuracy (before and after correction).

**(c)** Effect of finger speed on character- and word-level accuracy.

**Figure 4: Results from preliminary evaluations of our (a-b) Stage 2 algorithms and (c) the effect of finger speed on overall character- and word-level accuracy.**

end of a word is visible, we sort the aggregated character classifications and accept the most frequent classification. This process can be improved by incorporating a language dictionary model, albeit at the expense of efficiency. A text-to-speech engine reads back the identified word.

To investigate the effect of finger movement speed on OCR accuracy, we recorded five different speeds using a single line of text. The results are presented in Figure 4c. With greater speed, motion blur is introduced, and feature tracking becomes less accurate. In our experience, a “natural” finger speed movement for sighted readers is roughly 2-3cm/s. So, with the current prototype, one must move slower than natural for good performance. We plan on compensating for this effect in the future using image stabilization and motion blur removal, as well as incorporating a higher frame rate camera (100fps).

## 4 User Study to Assess Audio and Haptic Feedback

Our current prototype implementation supports haptic and audio feedback, but how best to implement this feedback for efficient direct-touch reading is an open question. Ultimately, we plan to conduct a holistic user evaluation of the system to assess the combined real-time OCR and finger guidance for a variety of reading tasks. At this stage, however, our goal was to refine the finger guidance component of the system by conducting a preliminary evaluation of three types of feedback: (1) audio only, (2) haptic only, and (3) a combined audio and haptic approach. We conducted a user study with four visually impaired participants to collect subjective and performance data on these three types of feedback. To isolate the finger guidance from the current OCR approach, we used a custom iPad app that simulates the experience of using the full system.

### 4.1 Method

**Participants.** We recruited four VI participants; details are shown in Table 1. All four participants had braille experience, and three reported regular use of screen readers.

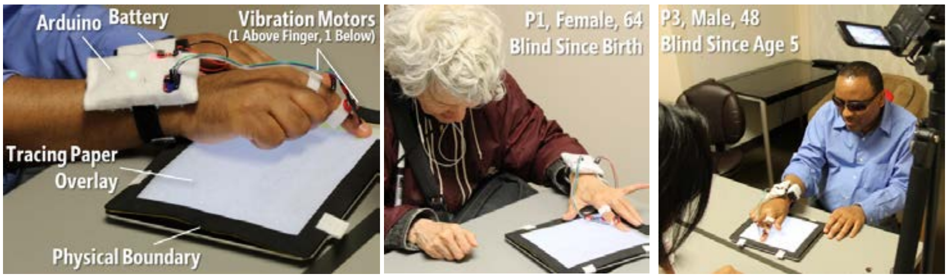
**Test apparatus.** The setup simulated the experience of reading a printed sheet of paper with HandSight (Figure 5). It consisted of the hand-mounted haptic component of the HandSight system controlled by an Arduino Micro, which was in turn connected via Bluetooth to an Apple iPad running a custom experimental app. The iPad was outfitted with a thin foam rectangle as a physical boundary around the edge of the screen to simulate the edge of a sheet of paper, and was further covered by a piece of tracing paper to provide the feel of real paper and to reduce friction. The app displayed text documents, guiding the user to trace each line of the document from left to right and top to bottom. As the user traced their finger on the screen, text-to-speech audio was generated, along with the following feedback guidance cues: start and end of a line of text, end of a paragraph, and vertical guidance for when the finger strayed above or below the current line. Lines were 36 pixels in height and vertical guidance began when the finger was more than 8 pixels above or below the vertical line center.

**Feedback conditions tested.** We compared three finger guidance options:

- *Audio only.* All guidance cues were provided through non-speech audio. The start and end of line cues were each a pair of tonal percussive (xylophone) notes played in ascending or descending order, respectively. The end of paragraph sound was a soft vibrophone note. When the user's finger drifted below or above a line, a continuous audio tone would be played to indicate that proper corrective movement. A lower tone (300 Hz) was played to indicate downward corrective movement (*i.e.*, the user was above the line). The pitch decreased at a rate of 0.83Hz/pixel to a minimum of 200Hz at 127 pixels above the line. A higher tone (500 Hz) was used to indicate upward corrective movement (up to a maximum of 600Hz with the same step value as before).
- *Haptic only.* The haptic feedback consists of two finger-mounted haptic motors, one on top and one underneath the index finger (see Section 3.2). Based on piloting within the research team, the motors were placed on separate finger segments (phalanges) so that the signal from each was easily distinguishable. To cue the start of a line, two short pulses played on both motors, with the second pulse more intense than the first; the reverse pattern indicated the end of a line. For the end of a paragraph, each motor vibrated one at a time, which repeated for a total of four pulses. For vertical guidance, when the finger strayed too high, the motor underneath the finger vibrated, with the vibration increasing in intensity from a

| ID | Age | Gender | Handed-ness | Level of Vision | Years of Vision Loss | Diagnosed Med. Condition | Hearing Difficulties |
|---|---|---|---|---|---|---|---|
| **P1** | 64 | Female | Left | Totally blind | Since birth | Retinopathy of prematurity | N/A |
| **P2** | 61 | Female | Left | Totally blind | Since birth | Retinopathy of prematurity | Slight hearing loss |
| **P3** | 48 | Male | Right | Totally blind | Since age 5 | N/A | N/A |
| **P4** | 43 | Female | Right | No vision one eye, 20/400 other eye | 30 years | Glaucoma | N/A |

**Table 1: Background of the four user study participants.**



(a) iPad test apparatus (b) Participant 1 (c) Participant 3

**Figure 5: User study setup and test apparatus: (a) overview; (b-c) in use by two participants..**

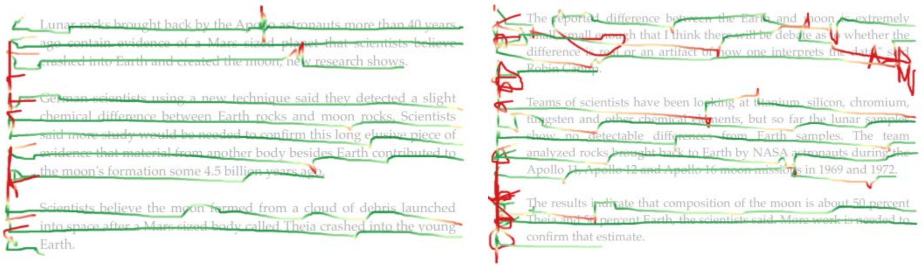
low perceivable value to maximum intensity, reached at 127 pixels above the line; below the line, the top motor vibrated instead (again with the maximum intensity reached at 127 pixels).

- *Combined audio-haptic.* This combined condition included all of the audio *and* haptic cues described above, allowing the two types of feedback to complement each other in case one was more salient for certain cues than the other.

**Procedure.** The procedure took up to 90 minutes. For each feedback condition, the process was as follows. First, we demonstrated the feedback cues for the start/end of each line, end of paragraph, and vertical guidance. Next, we loaded a training article and guided the user through the first few lines. Participants then finished reading the training article at their own pace. Finally, a test article was loaded and participants were asked to read through the text as quickly and accurately as possible. While we provided manual guidance as necessary to help participants read the *training* article (*e.g.*, adjusting their finger), no manual guidance was given during the *test* task. Four articles of approximately equivalent complexity were selected from *Voice of America* (a news organization), one for the training task and one to test each feedback condition; all articles had three paragraphs and on average 11.0 lines (*SD*=1.0) and 107.0 words (*SD*=13.5). The order of presentation for the feedback conditions was randomized per participant, while the test articles were always shown in the same order. Questions on ease of use were asked after each condition and at the end of the study. Sessions were video recorded, and all touch events were logged.

## 4.2 Findings

We analyzed subjective responses to the feedback conditions, and user performance based on logged touch events. Figure 6 shows a sample visualization from one participant (P1) completing the reading task in the *audio-only* and *haptic-only* conditions. Due to the small sample size, all findings in this section should be considered preliminary, but point to the potential impacts of HandSight and tradeoffs of different feedback.



Lunar rocks brought back by the Apollo astronauts more than 40 years ago contain evidence of a Mars-sized planet that scientists believe crashed into Earth and created the moon, new research shows.

German scientists using a new technique said they detected a slight chemical difference between Earth rocks and moon rocks. Scientists said more study would be needed to confirm this long elusive piece of evidence that material from another body besides Earth contributed to the moon's formation some 4.5 billion years ago.

Scientists believe the moon formed from a cloud of debris launched into space after a Mars-sized body called Theia crashed into the young Earth.

The reported difference between the Earth and moon is extremely small, small enough that I think there will be debate as to whether the difference is real or an artifact of how one interprets the data," said Robin Canup.

Teams of scientists have been looking at titanium, silicon, chromium, tungsten and other chemical elements, but so far the lunar samples show no detectable differences from Earth samples. The team analyzed rocks brought back to Earth by NASA astronauts during the Apollo 11, Apollo 12 and Apollo 16 moon missions in 1969 and 1972.

The results indicate that composition of the moon is about 50 percent Theia and 50 percent Earth, the scientists said. More work is needed to confirm that estimate.

(a) Participant 1 finger trace (audio only condition)

(b) Participant 1 finger trace (haptic only condition)

**Figure 6: Our iPad test apparatus allowed us to precisely track and measure finger movement. Example trace graphs for Participant 1 (P1) across the audio- and haptic-only conditions are shown above (green is on line; red indicates off-line and guidance provided). These traces were also used to calculate a range of performance measures. For example, the average overall time to read a line for P1 took 11.3s (*SD*=3.9s) in the audio condition and 18.9s (*SD*=8.3s) in the haptic condition. The average time to find the beginning of the next line (traces not shown above for simplicity but were recorded) was 2.2s (*SD*=0.88s) in the audio condition and 2.7s (*SD*=2.4s) in the haptic condition.**

In terms of overall preference, three participants preferred *audio-only* feedback; see Table 2. Reasons included that they were more familiar with audio than haptic signals (P1, P3), and that it was easier to attend to text-to-speech plus audio than to text-to-speech plus haptic (P4). P2's most preferred condition was the *combined* feedback, because she liked to have audio cues for line tracing and haptic cues for start/end of line notifications. In contrast, *haptic-only* feedback was least preferred by three participants. For example, concerned by the desensitization of her nerves, P1 expressed that: *"…if your hands are cold, a real cold air-conditioned room, it's [my tactile sensation] not going to pick it up as well."* P4 also commented on being attuned to sound even in the haptic condition: *"You don't know if it's the top or the bottom [vibrating]…It was the same noise, the same sound."* As shown in Figure 7, ease of use ratings on specific components of the task mirrored overall preference rankings.

| | Rank 1 | Rank 2 | Rank 3 |
|---|---|---|---|
| **P1** | Audio | Combined | Haptic |
| **P2** | Combined | Audio | Haptic |
| **P3** | Audio | Haptic | Combined |
| **P4** | Audio | Combined | Haptic |

**Table 2: Overall preference rankings per participant. Audio feedback was the most positively received.**

| | Braille | Screen Reader | Printed Text |
|---|---|---|---|
| **P1** | 3 | 3 | 3 |
| **P2** | 3 | 5 | 5 |
| **P3** | 4 | 4 | 4 |
| **P4** | 5 | 5 | 5 |

**Table 3: Ratings comparing prior text reading experiences with HandSight; 1-much worse to 5-much better.**

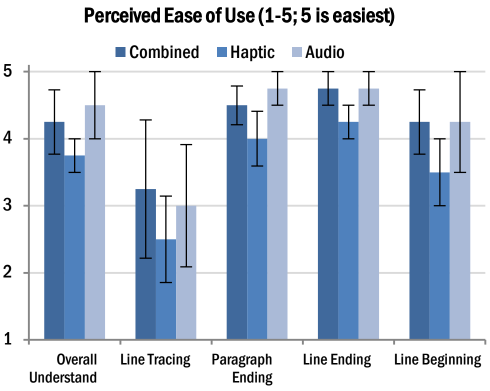


**Figure 7: Avg perceived ease of use of different text guidance attributes based on a 5-point scale (1-*very difficult*; 5-*very easy*). Error bars are stderr (*N*=4).**

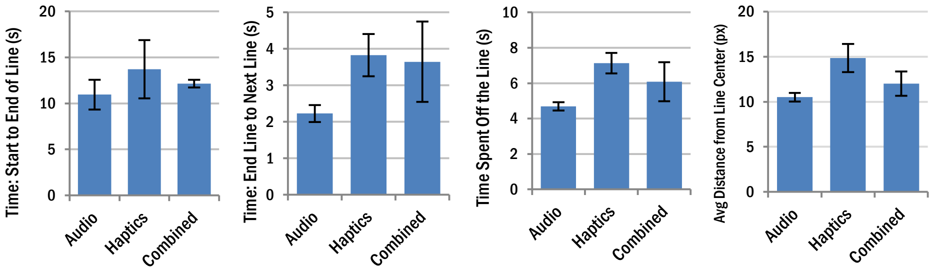


**Figure 8: Average performance data from the four user study participants across the three feedback conditions. While preliminary, these results suggest that audio-only feedback may be more effective than the other options tested. Error bars show standard error; (*N*=4).**

Participants were also asked to compare their experience with HandSight to braille, screen readers and printed-text reading using 5-point scales (*1-much worse* to *5-much better*). As shown in Table 3, HandSight was perceived to be at least as good (3) or better compared to each of the other reading activities. In general, all participants appreciated HandSight because it allowed them to become more independent when reading non-braille printed documents. For example, P3 stated, *"It puts the blind reading on equal footing with rest of the society, because I am reading from the same reading material that others read, not just braille, which is limited to blind people only"*. P1, who had experience with Optacon [26], Sara CE, and other printed-text scanning devices also commented on HandSight's relative portability.

In terms of performance, we examined four primary measures averaged across all lines per participant (Figure 8): average absolute vertical distance from the line center, time spent off the line (*i.e.*, during which vertical feedback was on), time from start to end of a line, and time from the end of a line to the start of the next line. While it is difficult to generalize based on performance data from only four participants, *audio-only* may offer a performance advantage over the other two conditions. *Audio-only* resulted in the lowest average vertical distance to the line center for all participants. Compared to the *haptic-only* condition, *audio-only* reduced the amount of time spent off the line by about half. It was also faster for all participants than *haptic-only* in moving from the end of a line to the start of the next line. A larger study is needed to confirm these findings and to better assess what impact the feedback conditions have on reading speed from start to end of a line.

## 5 Discussion

Though preliminary, our research contributes to the growing literature on wearables to improve access to the physical world for the blind (*e.g.*, [12], [19], [21]). The design and initial algorithmic evaluation of our current HandSight prototype show the feasibility of our approach, and highlight important technical issues that we must consider. Additionally, our user study, which evaluated three finger-guidance approaches using a controlled setup (the iPad test apparatus), found that, in contrast to prior work [21], haptic feedback was the *least* preferred guidance condition. The pitch-controlled audio feedback condition was not only subjectively rated the most preferred but also appeared to improve user performance. Clearly, however, more work is needed. Below, we discuss our preliminary findings and opportunities for future work.

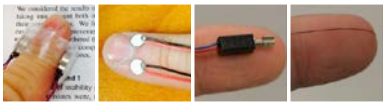


**Figure 9: We are evaluating a range of micro-haptic actuators: (a) 10×2.7mm$^2$ vibro-discs; (b) 5×0.4 mm$^2$ piezo discs; (c) 3×8 mm$^2$ vibro-motors; (d) 0.08mm Flexinol wire (shape memory alloy).**

**Haptic Feedback.** Though we have created many different types of finger-mounted haptic feedback in our lab, we tested only one in the user study: when the user moved above or below the current line, s/he would feel a continuous vibration proportional in strength to the distance from the vertical line center. We plan to experiment with form factors, haptic patterns (*e.g.*, intensity, frequency, rhythm, pressure), number of haptic devices on the finger, as well as the type of actuator itself (Figure 9). While our current haptic implementation performed the worst of the feedback conditions, we expect that, ultimately, some form of haptics will be necessary for notifications and finger guidance.

**Blind reading.** Compared to current state-of-the-art reading approaches, our long-term goals are to: (1) provide more intuitive and precise control over scanning and text-to-speech; (2) increase spatial understanding of the text layout; and (3) mitigate camera framing, focus, and lighting issues. Moreover, because pointing and reading are tightly coupled, finger-based interaction intrinsically supports advanced features such as rereading (for sighted readers, rereading occurs 10-15% of the time [63] and increases comprehension and retainment [64], [65]). We focused purely on reading simple document text, but we plan to investigate more complex layouts so the user can sweep their finger over a document and sense where pictures are located, headings, and so on. We will explore a variety of documents (*e.g.*, plain text, magazines, bills) and household objects (*e.g.*, cans of food, cleaning supplies), and examine questions such as: How should feedback be provided to indicate where text/images are located? How should advanced features such as re-reading, excerpting, and annotating be supported, perhaps, through additional gestural input and voice notes?

**Computer Vision.** Our preliminary algorithms are efficient and reasonably accurate, but there is much room for improvement. By incorporating constraints on lower-level text features we may be able to rectify vertical perspective effects and affine skew. We can also apply deblurring and image stabilization algorithms to improve the maximum reading speed the system is able to support. Robust and efficient document mosaicking and incorporation of prior knowledge will likely be a key component for supporting a wider range of reading tasks.

**Multi-sensory approach.** Currently, our prototype relies on only local information gleaned from the on-finger camera. However, in the future, we would like to combine camera streams from both a body-mounted camera (*e.g.,* Orcam [12]) and a finger-mounted camera. We expect the former could provide more global, holistic information about a scene or text which could be used to guide the finger towards a target of interest or to explore the physical document's layout. We could also use the information to improve the performance of the OCR algorithms, by dynamically training the classifier on the page fonts and creating a generative model (*e.g.,* [66]).

## 6 Conclusion

Our overarching vision is to transform how people with VI access visual information through touch. Though we focused specifically on reading, this workshop paper offers a first step toward providing a general platform for touch-vision applications.

## 7 References

[1] D. Pascolini and S. P. Mariotti, "Global estimates of visual impairment: 2010.," *Br. J. Ophthalmol.*, vol. 96, no. 5, pp. 614–8, May 2011.

[2] National Eye Institute at the National Institute of Health, "Blindness Statistics and Data." [Online]. Available: http://www.nei.nih.gov/eyedata/blind.asp. [Accessed: 10-Mar-2014].

[3] D. Dakopoulos and N. G. Bourbakis, "Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey," *IEEE Trans. Syst. Man, Cybern. Part C (Applications Rev.*, vol. 40, no. 1, pp. 25–35, Jan. 2010.

[4] G. Balakrishnan, G. Sainarayanan, R. Nagarajan, and S. Yaacob, "Wearable real-time stereo vision for the visually impaired," *Eng. Lett.*, vol. 14, no. 2, pp. 6–14, 2007.

[5] J. A. Hesch and S. I. Roumeliotis, "Design and Analysis of a Portable Indoor Localization Aid for the Visually Impaired," *Int. J. Rob. Res.*, vol. 29, no. 11, pp. 1400–1415, Jun. 2010.

[6] A. Hub, J. Diepstraten, and T. Ertl, "Design and Development of an Indoor Navigation and Object Identification System for the Blind," *SIGACCESS Access. Comput.*, no. 77–78, pp. 147–152, Sep. 2003.

[7] R. Manduchi, "Mobile Vision as Assistive Technology for the Blind: An Experimental Study," in *Computers Helping People with Special Needs SE - 2*, vol. 7383, K. Miesenberger, A. Karshmer, P. Penaz, and W. Zagler, Eds. Springer Berlin Heidelberg, 2012, pp. 9–16.

[8] A. Helal, S. E. Moore, and B. Ramachandran, "Drishti: an integrated navigation system for visually impaired and disabled," in *Proceedings Fifth International Symposium on Wearable Computers*, 2001, pp. 149–156.

[9] S. Krishna, G. Little, J. Black, and S. Panchanathan, "A Wearable Face Recognition System for Individuals with Visual Impairments," in *Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility*, 2005, pp. 106–113.

[10] S. Krishna, D. Colbry, J. Black, V. Balasubramanian, and S. Panchanathan, "A Systematic Requirements Analysis and Development of an Assistive Device to Enhance the Social

Interaction of People Who are Blind or Visually,” in *Impaired,” Workshop on Computer Vision Applications for the Visually Impaired (CVAVI 08), European Conference on Computer Vision ECCV 2008*, 2008.

[11] L. Gade, S. Krishna, and S. Panchanathan, “Person Localization Using a Wearable Camera Towards Enhancing Social Interactions for Individuals with Visual Impairment,” in *Proceedings of the 1st ACM SIGMM International Workshop on Media Studies and Implementations That Help Improving Access to Disabled Users*, 2009, pp. 53–62.

[12] OrCam Technologies Ltd, “OrCam - See for Yourself.” [Online]. Available: http://www.orcam.com/. [Accessed: 23-Jun-2014].

[13] F. Iannacci, E. Turnquist, D. Avrahami, and S. N. Patel, “The Haptic Laser: Multi-sensation Tactile Feedback for At-a-distance Physical Space Perception and Interaction,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011, pp. 2047–2050.

[14] V. Khambadkar and E. Folmer, “GIST: A Gestural Interface for Remote Nonvisual Spatial Perception,” in *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology*, 2013, pp. 301–310.

[15] A. Israr, O. Bau, S.-C. Kim, and I. Poupyrev, “Tactile Feedback on Flat Surfaces for the Visually Impaired,” in *CHI ’12 Extended Abstracts on Human Factors in Computing Systems*, 2012, pp. 1571–1576.

[16] J. F. Norman and A. N. Bartholomew, “Blindness enhances tactile acuity and haptic 3-D shape discrimination.,” *Atten. Percept. Psychophys.*, vol. 73, no. 7, pp. 2323–31, Oct. 2011.

[17] D. Goldreich and I. M. Kanics, “Tactile Acuity is Enhanced in Blindness,” *J. Neurosci.*, vol. 23, no. 8, pp. 3439–3445, Apr. 2003.

[18] S. Karim, A. Andjomshoaa, and A. M. Tjoa, “Exploiting SenseCam for Helping the Blind in Business Negotiations,” in *Computers Helping People with Special Needs SE - 166*, vol. 4061, K. Miesenberger, J. Klaus, W. Zagler, and A. Karshmer, Eds. Springer Berlin Heidelberg, 2006, pp. 1147–1154.

[19] S. Nanayakkara, R. Shilkrot, K. P. Yeo, and P. Maes, “EyeRing: A Finger-worn Input Device for Seamless Interactions with Our Surroundings,” in *Proceedings of the 4th Augmented Human International Conference*, 2013, pp. 13–20.

[20] X.-D. Yang, T. Grossman, D. Wigdor, and G. Fitzmaurice, “Magic Finger: Always-available Input Through Finger Instrumentation,” in *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*, 2012, pp. 147–156.

[21] R. Shilkrot, J. Huber, C. Liu, P. Maes, and N. S. Chandima, “FingerReader: A Wearable Device to Support Text Reading on the Go,” *CHI ’14 Ext. Abstr. Hum. Factors Comput. Syst.*, no. Vi, pp. 2359–2364, 2014.

[22] F. S. Cooper, J. H. Gaitenby, and P. W. Nye, *Evolution of reading machines for the blind: Haskins Laboratories’ research as a case history*. Haskins Laboratories, 1983.

[23] M. Capp and P. Picton, “The optophone: an electronic blind aid,” *Eng. Sci. Educ. J.*, vol. 9, no. 3, pp. 137–143, 2000.

[24] E. F. D’Albe, “On a Type-Reading Optophone,” *Proc. R. Soc. London. Ser. A*, vol. 90, no. 619, pp. 373–375, 1914.

[25] J. C. Bliss, “A Relatively High-Resolution Reading Aid for the Blind,” *Man-Machine Syst. IEEE Trans.*, vol. 10, no. 1, pp. 1–9, Mar. 1969.

[26] D. Kendrick, “From Optacon to Oblivion: The Telesensory Story,” *American Foundation for the Blind AccessWorld Magazine*, vol. 6, no. 4, 2005.

[27] Intel, “Intel Reader.” [Online]. Available: http://www.intel.com/pressroom/kits/healthcare/reader/. [Accessed: 10-Jan-2014].

[28] knfb Reading Technology Inc., “knfb Reader Classic.” [Online]. Available: http://www.knfbreader.com/products-classic.php. [Accessed: 22-Jun-2014].

[29] V. Gaudissart, S. Ferreira, C. Thillou, and B. Gosselin, “SYPOLE: mobile reading assistant for blind people,” in *9th Conference Speech and Computer (SPECOM)*, 2004.

[30] Blindsight, “Text Detective.” [Online]. Available: http://blindsight.com/textdetective/. [Accessed: 01-Nov-2013].

[31] knfb Reading Technology Inc., “kReader Mobile.” [Online]. Available: http://www.knfbreader.com/products-kreader-mobile.php. [Accessed: 24-Jun-2014].

[32] S. Mori, C. Y. Suen, and K. Yamamoto, “Historical review of OCR research and development,” *Proc. IEEE*, vol. 80, no. 7, pp. 1029–1058, Jul. 1992.

[33] H. Shen and J. Coughlan, “Towards a Real-Time System for Finding and Reading Signs for Visually Impaired Users,” in *Computers Helping People with Special Needs SE - 7*, vol. 7383, K. Miesenberger, A. Karshmer, P. Penaz, and W. Zagler, Eds. Springer Berlin Heidelberg, 2012, pp. 41–47.

[34] X. Chen and A. L. Yuille, “Detecting and reading text in natural scenes,” in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, 2004, vol. 2, pp. II–366–II–373 Vol.2.

[35] K. Wang, B. Babenko, and S. Belongie, “End-to-end scene text recognition,” in *Computer Vision (ICCV), 2011 IEEE International Conference on*, 2011, pp. 1457–1464.

[36] C. Jayant, H. Ji, S. White, and J. P. Bigham, “Supporting Blind Photography,” in *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility*, 2011, pp. 203–210.

[37] R. Manduchi and J. M. Coughlan, “The last meter: blind visual guidance to a target,” in *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems (CHI’14)*, 2014, p. To appear.

[38] S. K. Kane, B. Frey, and J. O. Wobbrock, “Access Lens: A Gesture-based Screen Reader for Real-world Documents,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2013, pp. 347–350.

[39] OrCam as uploaded by YouTube user Amnon Shashua, “OrCam at Digital-Life-Design (DLD) in Munich,” *Digital-Life-Design*, 2014. [Online]. Available: http://youtu.be/3m9ivtJI6iA?t=2m10s. [Accessed: 22-Jun-2014].

[40] OrCam as uploaded by YouTube user Amnon Shashua, “OrCam TED@NYC,” *TED@NYC*, 2013. [Online]. Available: http://youtu.be/_3XVsCsscyw?t=4m53s. [Accessed: 22-Jun-2014].

[41] A. Crossan and S. Brewster, “Multimodal Trajectory Playback for Teaching Shape Information and Trajectories to Visually Impaired Computer Users,” *ACM Trans. Access. Comput.*, vol. 1, no. 2, pp. 12:1–12:34, Oct. 2008.

[42] B. Plimmer, P. Reid, R. Blagojevic, A. Crossan, and S. Brewster, “Signing on the Tactile Line: A Multimodal System for Teaching Handwriting to Blind Children,” *ACM Trans. Comput. Interact.*, vol. 18, no. 3, pp. 17:1–17:29, Aug. 2011.

[43] J. Su, A. Rosenzweig, A. Goel, E. de Lara, and K. N. Truong, “Timbremap: Enabling the Visually-impaired to Use Maps on Touch-enabled Devices,” in *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services*, 2010, pp. 17–26.

[44] S. Harada, H. Takagi, and C. Asakawa, “On the Audio Representation of Radial Direction,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011, pp. 2779–2788.

[45] K. Yatani and K. N. Truong, “SemFeel: A User Interface with Semantic Tactile Feedback for Mobile Touch-screen Devices,” in *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology*, 2009, pp. 111–120.

[46] K. Yatani, N. Banovic, and K. Truong, “SpaceSense: Representing Geographical Information to Visually Impaired People Using Spatial Tactile Feedback,” in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2012, pp. 415–424.

[47] N. Noble and B. Martin, “Shape discovering using tactile guidance,” in *Proceeding of the 6th International Conference EuroHaptics*, 2006.

[48] U. Oh, S. K. Kane, and L. Findlater, “Follow that sound: using sonification and corrective verbal feedback to teach touchscreen gestures,” in *Proceedings of the ACM SIGACCESS International Conference on Computers and Accessibility (ASSETS 2013)*, 2013, p. To appear.

[49] S. J. Lederman and R. L. Klatzky, “Hand movements: A window into haptic object recognition,” *Cogn. Psychol.*, vol. 19, no. 3, pp. 342–368, 1987.

[50] K. Johnson, “Neural basis of haptic perception,” in *Stevens’ Handbook of Experimental Psychology: Volume 1: Sensation and Perception*, 3rd Editio., H. Pashler and S. Yantis, Eds. Wiley Online Library, 2002, pp. 537–580.

[51] T. S. Saponas, D. S. Tan, D. Morris, R. Balakrishnan, J. Turner, and J. A. Landay, “Enabling Always-available Input with Muscle-computer Interfaces,” in *Proceedings of the 22Nd Annual ACM Symposium on User Interface Software and Technology*, 2009, pp. 167–176.

[52] D. Morris, T. S. Saponas, and D. Tan, “Emerging input technologies for always-available mobile interaction,” *Found. Trends Human--Computer Interact.*, vol. 4, no. 4, pp. 245–316, 2010.

[53] T. S. Saponas, “Supporting Everyday Activities through Always-Available Mobile Computing,” University of Washington, 2010.

[54] AWAIBA, “NanEye Medical Image Sensors.” [Online]. Available: http://www.awaiba.com/en/products/medical-image-sensors/. [Accessed: 10-Jan-2014].

[55] Samsung, “Samsung Galaxy Gear.” [Online]. Available: http://www.samsung.com/us/mobile/wearable-tech/SM-V7000ZKAXAR. [Accessed: 10-Jan-2014].

[56] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.

[57] L. Jagannathan and C. Jawahar, “Perspective Correction Methods for Camera Based Document Analysis,” *Proc. First Int. Work. Camera-based Doc. Anal. Recognit.*, pp. 148–154, 2005.

[58] V. Zaliva, “Horizontal Perspective Correction in Text Images,” 2012. [Online]. Available: http://notbrainsurgery.livejournal.com/40465.html.

[59] I. S. Abuhaiba, “Efficient OCR using Simple Features and Decision Trees with Backtracking,” *Arab. J. Sci. Eng.*, 2006, vol. 31, no. 2, pp. 223–244, 2006.

[60] E. Rosten and T. Drummond, “Fusing points and lines for high performance tracking,” in *Tenth IEEE International Conference on Computer Vision (ICCV’05) Volume 1*, 2005, vol. 2, pp. 1508–1515 Vol. 2.

[61] C. Tomasi and T. Kanade, *Detection and Tracking of Point Features*. Carnegie Mellon University Technical Report CMU-CS-91-132, 1991.

[62] M. A. Fischler and R. C. Bolles, “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography,” *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[63] R. Keefer, Y. Liu, and N. Bourbakis, “The Development and Evaluation of an Eyes-Free Interaction Model for Mobile Reading Devices,” *Human-Machine Syst. IEEE Trans.*, vol. 43, no. 1, pp. 76–91, 2013.

[64] S. L. Dowhower, “Repeated Reading: Research into Practice,” *Read. Teach.*, vol. 42, no. 7, pp. pp. 502–507, 1989.

[65] B. A. Levy, “Text processing: Memory representations mediate fluent reading,” *Perspect. Hum. Mem. Cogn. aging Essays honour Fergus Craik*, pp. 83–98, 2001.

[66] J. Lucke, “Autonomous cleaning of corrupted scanned documents — A generative modeling approach,” in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3338–3345.