Third Eye: A Shopping Assistant for the Visually Impaired

Peter A. Zientara, Sooyeon Lee, and Gus H. Smith, Pennsylvania State University
Rorry Brenner and Laurent Itti, University of Southern California
Mary B. Rosson and John M. Carroll, Pennsylvania State University
Kevin M. Irick, SiliconScapes
Vijaykrishnan Narayanan, Pennsylvania State University

Through a combination of wearable cameras, hardware accelerators, and algorithms, a vision-based automatic shopping assistant allows users with limited or no sight to select products from grocery shelves.

Globally, more than 285 million people are visually impaired; for them, tasks that are deemed trivial by those with normal sight—such as picking up a dropped object—are a major undertaking. In recent years, breakthroughs in visual sensing technology have stimulated efforts to help persons with visual impairment (PVIs) become more independent. Signal-processing technologies, central to visual augmentation, help PVIs navigate inside and outdoors[1] and identify objects—activities that are critical in everyday tasks such as shopping, particularly in supermarkets. Even sighted people are often overwhelmed by the amount and variety of products in a typical US grocery store, which can have 35,000 unique items presented in as many as 30 aisles across 45,000 square meters. For PVIs, grocery shopping is like navigating an area the size of a football field with moving carts, people, and displays at every turn.

To accomplish this daunting task, some PVIs turn to family, friends, or community volunteers. Others must rely on store assistants, who can make the shopping experience enjoyable or unpleasant, depending on their knowledge and personality. Online shopping and home-delivery services attempt to mitigate the challenges

 0018-9162/17/$33.00 © 2017 IEEE

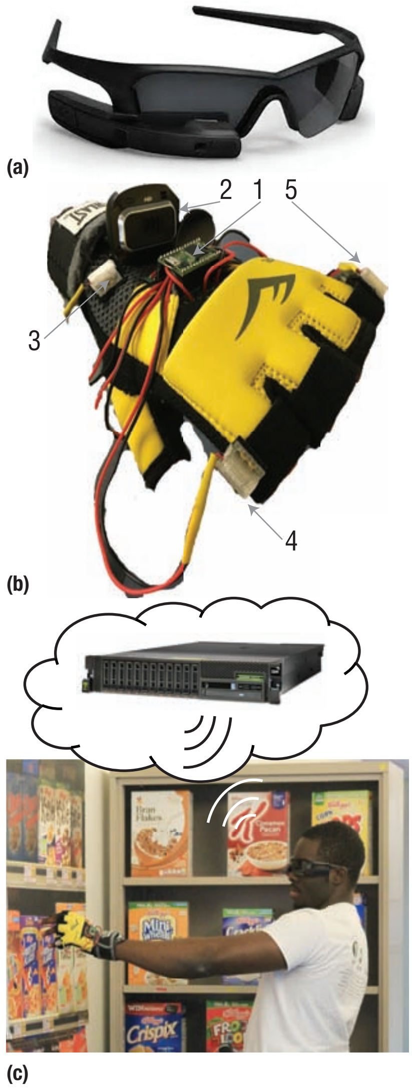


FIGURE 1. How Third Eye works. In this illustrative scenario, cameras in (a) smart glasses and (b) a glove capture video that is streamed wirelessly to (c) a server for video analysis. The system uses the results to guide the user through either audio commands based on images from the glasses to navigate toward the desired item's proximity or vibrations in the glove to acquire the item. The vibrations occur in different areas, depending on which direction the user's hand needs to move to acquire the desired product. The glove consists of six parts: (1) a control and communication module, (2) the camera, (3) an up vibration motor, (4) a right vibration motor, and (5) a left vibration motor. Not visible is the down vibration motor, which is on the underside of the palm.

of grocery shopping for PVIs by eliminating trips to the store,[2,3] but these methods limit both social interaction and product choices. Assistive technologies, another mitigation effort, are basically computer vision algorithms running on mobile platforms. Their rudimentary support, which is limited to verifying one already located object, rapidly breaks down in environments where multiple similar items are in tightly packed configurations. Even advanced vision algorithms often fail to distinguish between two variants of the same cereal brand, for example. Unique segmentation—essential to recognizing one item in a tightly packed configuration—requires brute-force, computationally intensive searches that are slow in responding to queries.

To address these problems and provide PVIs with a more personally satisfying in-store experience, we developed a vision-based automatic shopping assistant that helps users select the desired item from a typical grocery store shelf. Our work is part of the NSF-funded Visual Cortex on Silicon project (www.cse.psu.edu/research/visualcortexonsilicon.expedition). Third Eye—so named for the Asian cultural reference to the third eye of wisdom and compassion—addresses the critical challenge of providing suitable user interfaces for navigating in a crowded and visually diverse environment. Our observations of and interviews with PVIs helped us design Third Eye's navigation interfaces to both augment and complement skills that individuals already have, such as the ability to identify an obstacle with a cane or walking stick.

To enrich navigation and object selection, Third Eye relies on multimodal tactile and audio interfaces as well as a system based on the human-in-the-loop concept, in which the system interacts with the user to make navigation and selection decisions. For example, the system can ask the user to change a wearable camera's orientation so that the camera has a more accurate perspective on which to base image classification.

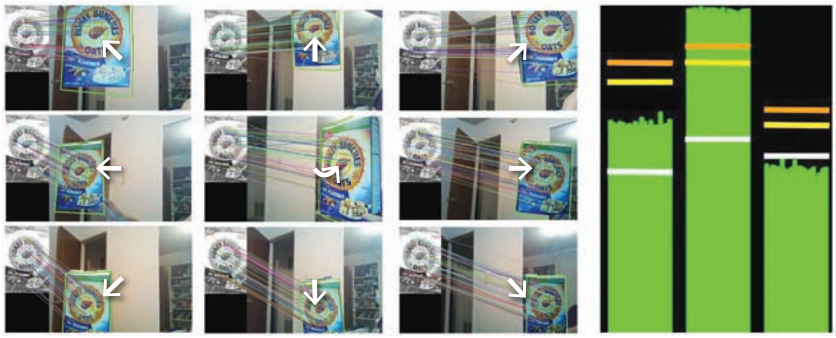
We have taken theq system through multiple evaluations and refinements based on actual PVI use, which provided insights into what constitutes an effective communication interface. Participants in our experiments commented favorably on the system, saying that, relative to barcode scanning, it felt more like "real shopping."

SYSTEM INFRASTRUCTURE

As Figure 1 shows, Third Eye consists of off-the-shelf smart glasses—equipped with a camera and audio channel—connected to a back-end server system that supports real-time video analytics. The camera is oriented to the user's right eye, and the smart glasses' field of view and resolution dictate the distance at which objects can be properly located and identified. Users also wear a glove with a camera on the hand used to grasp a product. The glove camera, which guides hand movements, must also orient the user to keep the product in view and avoid any occlusion from the arm used to point to and grasp it.

Third Eye uses the store's wireless infrastructure to send the videostream to the server, where computer vision algorithms analyze it and send back results. This feedback then becomes the basis for providing either audio commands or tactile vibration patterns that guide the user's steps and hands toward the desired item. To avoid latency between image capture and feedback signals, Third Eye incorporates a local wireless routing infrastructure, which ensures robust, low-latency communication with the server, and customized hardware to accelerate the vision algorithms' computations.

Images from the smart glasses are the basis for directions to move the user closer to the desired item. Images



(a) (b)

FIGURE 2. Version A of the Speeded up Robust Features (SURF) algorithm. (a) Directional commands (denoted by white arrows) through an audio channel guide the user so that the camera in the glasses can obtain at least a partial view of the product for matching. (b) Graph showing confidence levels for three product templates. Third Eye matches these points against points in the image template and assigns each match a confidence level, which it uses to decide if this is the desired product or if the user should move to another item. The horizontal colored lines represent threshold confidence levels.

from the glove camera then help orient the user to a view that provides enough information to the vision algorithms. Inspired by prior efforts that used sensor-equipped haptic gloves for visual search and interaction with smart glasses,[4,5] we incorporated vibrational feedback for hand movements into the user's glove. The glove has four micromotors to direct movement (up, down, left, and right), and simultaneous vibration of all four motors directs the hand forward to grasp a product that is in position. Although audible feedback has been used to provide directional navigation, on the basis of observing PVIs and their interactions in store environments, we opted to supplement it with tactile feedback in the form of vibrations. We found that it is not effective in noisy environments and draws unwanted attention from other shoppers. A headset might address these challenges but it would have to be highly specialized, incorporating bone conduction or the superimposition of instructions while allowing for environmental noise—a key navigational aid for PVIs.

VISION ALGORITHMS

Third Eye's vision algorithms assist users from the time they enter the store to the time they leave it. The first task is to identify the correct aisle. When users are in front of the products in the aisles, the camera captures the image and uses a gist-of-the-scene algorithm to localize their position.[6] The algorithm extracts a low-dimensional signature of the entire image that can support scene classification. For each videoframe from the camera, it aggregates statistics about the scene, including color distribution, luminance, and oriented edges. The aggregated statistics summarize the scene contents and translate it to a gist vector—a low-dimensional holistic feature vector—which a support vector machine (SVM) classifier processes to produce a category label for that scene, such as a cereal or coffee aisle.

If users are not in front of a desired aisle, they move to another aisle and repeat the process until they arrive at the desired aisle. Navigation between the aisles is not yet completely automated, but our system effectively integrates with the navigation skills PVIs already have. Once the user is in the right aisle, Third Eye uses Speeded up Robust Features (SURF), a feature-extraction algorithm and template-matching process that uses a known pattern of a named object to match an object in a partial view.[7] Third Eye extends SURF by providing a confidence percentage that a target item in partial view is the desired product. The score helps guide the camera position to obtain a better view of the object, increasing the confidence that the item being viewed is the desired product. The extended template-matching algorithm in Third Eye matches key points in the camera image with the template image. Each key-point match provides a location of the point in the new image along with a confidence of the match. Figure 2 shows sample results of matching with our first version of SURF, version A, which we subsequently refined to version B. In Figure 2a, Third Eye guides the user to a position that has at least a partial product view, which is required for SURF to begin the feature-extraction and matching process.

Confidence matching

Most confidence-generating algorithms use a fixed threshold to decide when the target item is in the captured image or select the item with the highest confidence level. Most algorithms that use a fixed confidence threshold will select an item in error because of the uncertainty introduced when the highest false positive outweighs the lowest true positive.[8] The highest false positive is the highest confidence calculated when the target item is *not* in the captured image. The lowest true positive is the lowest confidence calculated when the target item *is* in the captured image. Whenever the highest false positive is stronger than the lowest true positive, there is a range of uncertainty, because the system cannot clearly determine whether the object is present or not. In our approach, no answer is provided to the PVI in this uncertain region; rather, the user is directed to move to improve the confidence level before the system attempts any classification.

In Figure 2b, for example, the white horizontal bar denotes the lowest true positive, and the yellow horizontal bar denotes the highest false positive. The distance from the yellow to the orange horizontal bars is a safety buffer beyond the highest false positive. The value must surpass the orange bar for Third Eye to conclude that the image captured is indeed the target item. Given these confidence boundaries, the left green bar is in the range of uncertainty, so Third Eye would instruct the user to get a better view. The center green bar exceeds the safety buffer, so Third Eye would conclude that the target item is in the image. The right green bar is below even the lowest true positive, so the system would ignore it and not issue any instructions.

With the range of uncertainty in Figure 2b, Third Eye could match key points between the template image and the camera frame and calculate a homography matrix that is based on the key points' relative positions. The matrix essentially represents the current item orientation (as it is now viewed) relative to the item's orientation in the template image, which was always front-facing and centered in our experiments. This relative orientation is used to instruct the user to obtain a better view.

Expanding matching to the shelf

In version A, the goal was to get the user closer to the template orientation view. When the camera captured images that were nearer to this view, confidences were higher. However, when multiple objects on the same shelf have similar features or when the user is not close enough to the shelf for the algorithm to extract a sufficient number of feature points from the viewed object, there is no partial match and no way to calculate a homography matrix to increase confidence. To address this problem, Third Eye uses shelfwide feature extraction and matching, which is based on the store's planogram (a diagram of how products are placed on shelves), to guide the user closer to the shelf section that contains the desired item.

THIRD EYE USES SHELFWIDE FEATURE EXTRACTION AND MATCHING TO GUIDE THE USER CLOSER TO THE SHELF SECTION THAT CONTAINS THE DESIRED ITEM.

Once it matches a shelf, Third Eye uses the planogram to locate desired items and then analyzes the image only at a shelf location that contains those items. For example, if a shopper wanted a particular brand of spaghetti, Third Eye would not analyze the entire rice and pasta aisle, only the shelf that contained pasta and only the part of the shelf that contained spaghetti. Narrowing analysis to a particular shelf section eliminates some of the problems in version A of our algorithm, in which shoppers had to wait for Third Eye to determine that images were irrelevant before hearing additional instructions. Rather, in version B of our algorithm, key-point matching is for the entire shelf section and the user can be as far away as 12 feet. Consequently, many more feature points are available in determining a match, which enables robust identification among multiple similar objects on the shelf.

Algorithmic handoff from shelf to item

Once the user is within arm's reach of the shelf, the system transfers the camera feed from the smart glasses to the glove; that is, from navigating steps to specifying hand movements. To execute this handoff, we first tried basing it on the relative size of objects in the camera view, which Third Eye used to determine the user's proximity to the shelf. At a certain distance, Third Eye stopped directing on the basis of the smart glasses and began to initiate hand movements on the basis of the glove camera.

We abandoned the auto handoff approach after observing that it was not robust to changes in head alignment and angles; for example,

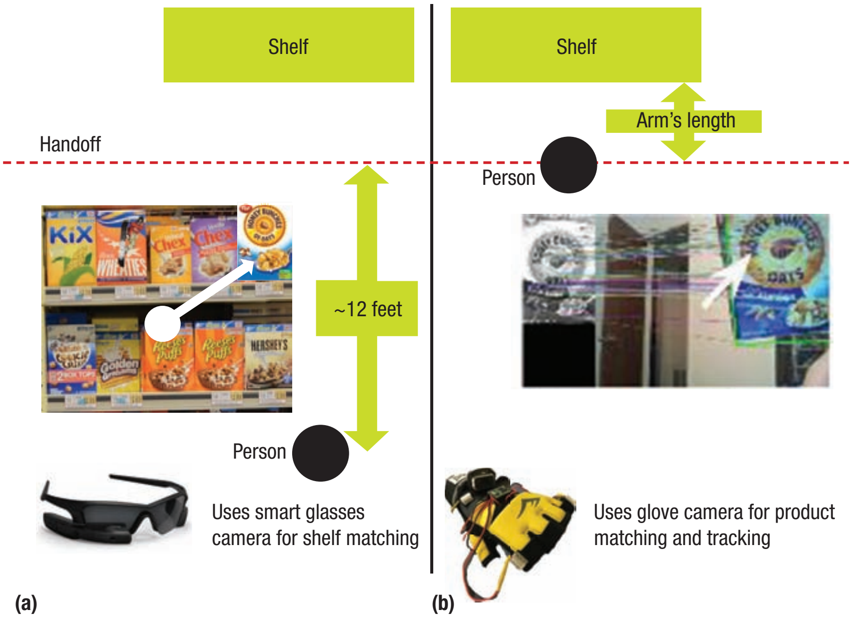


FIGURE 3. User-initiated handoff from shelf matching with smart glasses to item matching with the glove. (a) Third Eye uses shelf matching and a priori information about the store's shelf contents to guide the user's steps toward the desired item. Shelf matching can take place when the user is as far away as 12 feet from the target item. (b) Once the user is within arm's reach of the item, Third Eye uses object matching to guide hand movement. When the cane touches the shelf, the user initiates the handoff by muting voice commands from Third Eye, which is a signal for the system to begin using the glove camera and guide hand movements through glove vibrations.

head orientation to the object varied according to the user's height. On the basis of an experiment with a volunteer PVI, we discovered that the user might actually be a more effective handoff initiator. The PVI—unlike the sighted blindfolded participants in our proof-of-concept study—routinely used a cane and thus never ran into the shelves and was able to consistently stop at arm's length from them. We observed that the individual knew where to stop even when Third Eye's extraction of depth information failed, which led us to modify the transition from glasses to glove from system-initiated to user-initiated, as shown in Figure 3.

SYSTEM REFINEMENTS

In the various iterations of Third Eye refinement, we identified two major problems: feedback latency, which we eventually solved, and power drain, which has been mitigated.

Feedback latency

Even when the algorithms were running on an IBM POWER8 server with 160 logical cores operating at 3.6GHz, for a 1920 × 1080 image of an entire aisle, our latencies were 375 ms with 9,000 feature points in the image. We reduced latency to 170 ms by limiting the camera's field of view to a single shelf, which has around 4,000 feature points. Despite this latency reduction, we observed that delayed feedback from the system along with sudden jerky movements by shoppers caused the cameras to lose the object. To mitigate this problem, we used the NVIDIA GPU K1200 to accelerate object detection (shelf or item), which reduced feedback latency to 110 ms—enough to provide a testbed for a single user. From the results of testing with the K1200, we designed an accelerator to run on field-programmable gate arrays (FPGAs) that further reduces latency to support more concurrent videostreams. We are in the process of evaluating our customized acceleration scheme.

Power drain

In Third Eye's earlier versions, the smart glasses had a limited battery life. At first, we were streaming raw video over the wireless channel to the server, which drained the battery and clogged the wireless bandwidth. We then attempted to run software-based video compression on the lightweight processor in the smart glasses. This strategy provided no relief because the compression algorithm execution was slow. We finally moved to a camera with built-in hardware acceleration for video compression, which helped extend battery life to about 20 minutes.

One possible solution is to conserve power by leveraging additional information from the smart glasses. Instead of using the power-hungry video feedback, Third Eye can use information from the accelerometers and gyroscope to track the shopper's motion. This approach would enable video capture to be limited to locations that Third Eye knows are relevant on the basis of the planogram information the store has provided. We have already applied techniques similar to those used in vision processing to decode activities from the accelerometer signals, such walking, turning right, and turning left.

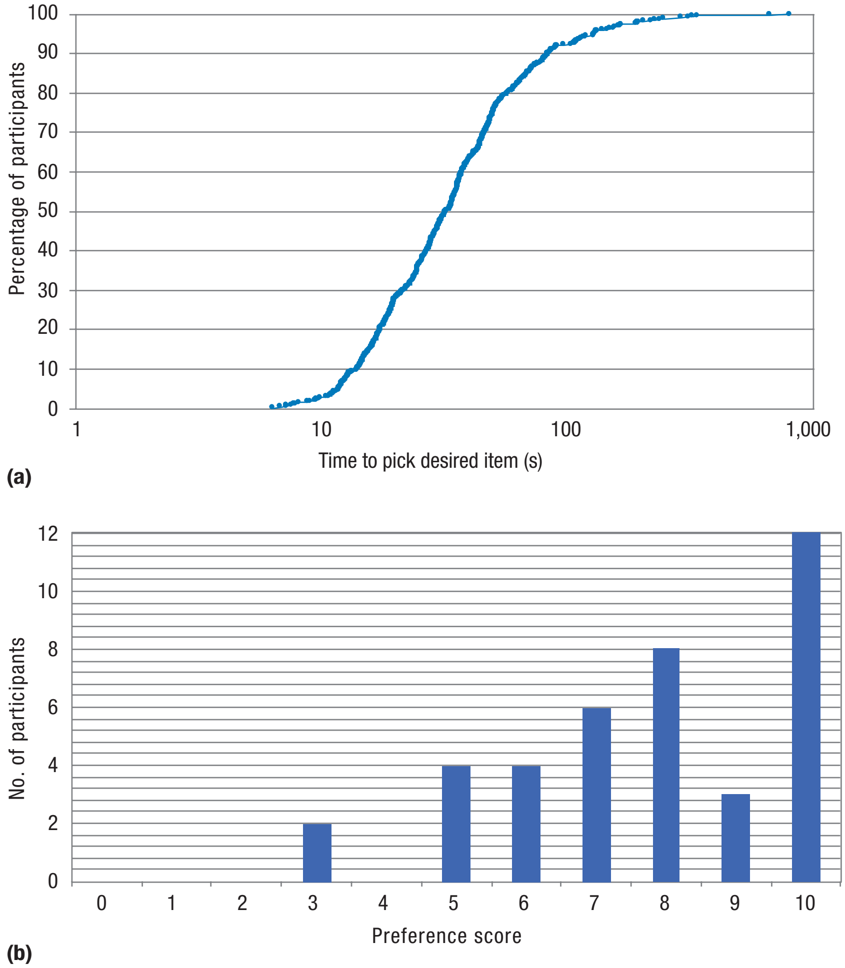


FIGURE 4. User preferences between Third Eye and barcode scanning in the proof-of-concept study. (a) Time for the user to grab the object after the first voice instruction from the system across 10 trials of the 42 participants. (b) Preference scores for the 39 users who responded. As the graph shows, 12 users rated Third Eye with a 10, indicating "prefer significantly more" over the off-the-shelf scanner in identifying target items. More than half the participants gave Third Eye a preference score of 7 or higher.

EVALUATION RESULTS

Beginning with a proof-of-concept study, we performed user experiments with Third Eye at different levels to refine it and explore additional enhancements. Even before the proof-of-concept evaluation with an integrated prototype, we conducted studies using a remote human viewer to carry out the work of the visual algorithm still under development (along the lines of a Wizard of Oz prototype). These earlier studies gave us a rich sampling of participant feedback and guidance systems.

Proof-of-concept study

We conducted the proof-of-concept study with 42 blindfolded sighted participants who used a glasses-mounted camera with speech guidance and version A of SURF (without shelf matching and the planogram information). For each trial, we positioned participants within 5 feet of the shelf with the target product. This position typically yielded enough feature points for each target to at least partially match an individual object, but sometimes participants had to move without instructions before Third Eye could find a partial match.

Procedure. As part of the study, we trained participants before trials in how to interpret speech guidance and to slightly reposition themselves, such as by turning, moving forward or backward, strafing (moving side to side), or crouching. In this way, Third Eye could register additional feature points. Once participants were positioned correctly, we asked them to reach out and grasp the identified targets. Separating the positioning and grasping phases ensured that the user's hand did not occlude the camera's view. We also found that speech guidance and self-repositioning were effective in eliminating false positive recognitions. This method worked successfully on each of the 10 trials conducted. As Figure 4a shows, the time for object pickup from first instruction ranged from 7.1 to 798.8 s with the majority of trials taking between 30 and 90 s.[9]

Comparison with barcode scanning. To provide a comparison point with standard off-the-shelf technology, we gave participants a barcode scanner with barcodes attached to the front-facing shelf below the target. For this activity, users stood close to the shelf and scanned the barcodes, without any guidance other than notification that scans were correct. All users successfully identified the target using the barcode scanner. However, as Figure 4b shows, 32 of 34 participants expressed a clear preference (a score higher than 6) for Third Eye's continuous guidance over the scanner. More than one participant noted that the system "is more like real shopping."

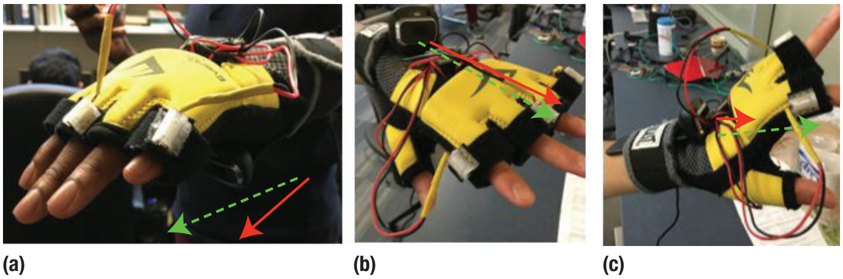


FIGURE 5. Glove-camera positioning. The red arrows depict the angle of the actual viewpoint for which directions are provided. The green dashed arrows represent the angle at which the user interprets commands. (a) The misalignment of angles causes the user to veer away from the object. (b) Angles are aligned correctly. (c) Angles are aligned but the object is occluded by the palm.

Auto handoff

After our proof-of-concept study, we evaluated the auto handoff we first used in version B of SURF. For this evaluation, we used blindfolded sighted users who were at least 10 feet away from the shelf and gave them only audio feedback and little a priori instruction. In most cases, the audio feedback system successfully brought the user in front of the desired object. In some cases, participants misinterpreted the directions, such as "move left," which they understood as turning left instead of moving left. Also, directions like "move left" or "move forward" did not specify how far, which led to significant variations in distance moved. On the basis of these results, we incorporated more precise directions, such as "move one step forward" or "move less than a step left."

Glove only

In another evaluation, we tested the camera system with only the glove camera and the vibrating motors to help with the final object grab when the user was within 3 feet of the target object. In this test, we used version A of SURF. Camera placement was a major challenge because we had to orient the camera in such a way that hand posture would not occlude the field of view—the main reference for providing movement instructions. In the prototype Third Eye, the camera was parallel to the arm to ensure that the hand and view were aligned. Unfortunately, in some experiments, the camera's orientation changed because of sudden jerky hand movements, which resulted in a lost field of view and consequent direction to the wrong objects. Figure 5 illustrates some of the glove design challenges we faced.

We also experimented with several configurations of the glove's vibrational feedback. After several positional trials, we arrived at a design that placed the micromotors on the fingers of the left hand to guide movement: the thumb to indicate a movement to right, the little finger to indicate a movement to left, one on top of the palm for up navigation, and one below the palm for down navigation. We continue to investigate the effectiveness of vibration intensities, durations, and patterns produced by motors for tactile messaging.[10]

Testing with a visually impaired user

In the previous tests, users were sighted but blindfolded, which we found does not accurately represent a PVI, who has had time to adjust to visual impairment. We only recently formatively assessed the Third Eye system with a PVI who has been working closely with our NSF project for the past two years, carrying out basic shopping interactions with a simulated store shelf. This individual had also participated in previous studies of a human-recognition prototype and field studies of PVI shopping practices. This experience made this individual an excellent feedback source in evaluating our functional prototype.

Auto handoff. We gave the PVI the smart glasses and only audio feedback with auto handoff (version B of SURF) but no vibrating glove. The PVI's initial position was about 10 feet away from a mock shelf containing nine products from the cereal aisle. The PVI then selected the object by moving through an audio menu in the glasses. After selecting the target object, the PVI was able to follow the audio feedback and navigate to a point close to the target using guidance from the speech feedback, which was based on images from the glasses camera. The speech feedback (for example, "move two steps slightly to the left") conveyed both direction and magnitude of movement, which enabled her to stay on the desired path.

Once the PVI was successfully positioned in front of the identified object, head alignment was not as steady as what we observed for blindfolded sighted users. We also noted that our estimate of the depth from PVI to shelf was not always reliable. Despite these drawbacks, the PVI consistently used Third Eye as an augmentation tool, stopping when her assistive cane touched the shelf.

User-initiated handoff. On the basis of the auto handoff test, we conducted another experiment with a user-initiated handoff, in which the PVI employed the cane in transitioning from the glasses camera and instructions for body movements to the glove camera and instructions for hand movements. This handoff occurred when the PVI's cane hit the bottom of the shelf, at which time the PVI directed Third Eye to mute audio commands based on the glasses camera. With this user-initiated handoff, the PVI was able to pick the desired object in the next trial. However, in other trials, the same

ABOUT THE AUTHORS

PETER A. ZIENTARA is a PhD student in computer science and engineering at Pennsylvania State University. His research interests include smart sensors, hardware acceleration, and embedded vision systems. Zientara received a BS in computer engineering from York College of Pennsylvania. Contact him at paz117@cse.psu.edu.

SOOYEON LEE is a PhD student in the College of Information Sciences and Technology at Pennsylvania State University. Her research interests include human–computer interaction for assistive technologies and community informatics. Lee received a BS in information sciences and technology from Pennsylvania State University. Contact her at sul131@psu.edu.

GUS H. SMITH is an MS student in computer science and engineering at Pennsylvania State University. His research interests include practical computer vision applications, hardware accelerator architectures, and emerging devices. Contact him at hfs5022@psu.edu.

RORRY BRENNER is a PhD student in neuroscience at the University of Southern California. His research interests include human-in-the-loop computer vision, assistive vision systems, object recognition, and artificial neural networks. Brenner received an MS in computer science from Carnegie Mellon University. Contact him at rorry.brenner@usc.edu.

LAURENT ITTI is a professor of computer science, psychology, and neurosciences at the University of Southern California. His current research interests include biologically inspired vision, particularly in visual attention, gist, saliency, and surprise, with applications to video compression, target detection, and robotics. Itti received a PhD in computation and neural systems from the California Institute of Technology. Contact him at itti@usc.edu.

MARY B. ROSSON is a professor and associate dean in the College of Information Sciences and Technology at Pennsylvania State University. Her research interests include scenario-based design of interactive systems, end-user programming and development, and online developmental communities. Rosson received a PhD in experimental psychology from the University of Texas at Austin. She is an ACM Distinguished Scientist. Contact her at mrosson@psu.edu.

JOHN M. CARROLL is a distinguished professor in the College of Information Sciences and Technology at Pennsylvania State University and director of the Center for Human–Computer Interaction. His research interests include methods and theory for human-centered design, particularly as applied to Internet tools for collaborative learning, problem solving, and creating and sustaining community. Carroll received a PhD in experimental psychology from Columbia University. He is a Fellow of IEEE and ACM. Contact him at jcarroll@ist.psu.edu.

KEVIN M. IRICK is CEO and founder of SiliconScapes. His research interests include high-performance, embedded, and cloud-based video analytics systems for security, surveillance, and retail applications; application-specific hardware accelerator design; hardware-assisted image processing and recognition; and high-performance computing on reconfigurable architectures. Irick received a PhD in computer science and engineering from Pennsylvania State University. Contact him at kevin.irick@siliconscapes.net.

VIJAYKRISHNAN NARAYANAN is a professor of computer science and engineering and of electrical engineering at Pennsylvania State University. His research interests include power-aware and reliable systems, embedded systems, nanoscale devices, interactions with system architectures, reconfigurable systems, computer architectures, networks-on-chip, and domain-specific computing. Narayanan received a PhD in computer science and engineering from the University of South Florida. He is a Fellow of IEEE and ACM. Contact him at vijay@cse.psu.edu.

PVI selected an object just below the desired object. These trials implied that objects positioned above the glasses camera's field of view were challenging and that we needed to further refine the directions provided.

Glove configurations. In the final test with the PVI, we evaluated glove configurations to determine the best camera orientation and whether feedback through vibrations or audio was more effective. The PVI seemed to prefer vibration over the audio feedback, pointing out that audio feedback without a headset draws unnecessary attention during shopping and interferes with the environmental sounds that offer cues to a PVI in orienting body position.

We also observed that the glove camera's orientation required more robust engineering. The lack of orientation reference caused the PVI to miss objects by less than a few inches in all trials. Moreover, when a miss occurred, the glove camera's field of view was occluded or the PVI was too close for the system to confidently identify the object. In these cases, the system stopped providing any directions for missed items because they exceeded the highest confidence score (see Figure 3b).

After 30 minutes of picking objects (including system pauses) with the user-initiated handoff and vibration and audio feedback, the PVI acquired the desired cereal box and placed it in the cart. Unfortunately, we had to return many other items to their shelves, as the PVI often selected the item immediately adjacent to the desired product.

Inspired by testing with the PVI, we are now integrating new modules to ensure that PVIs spend more time putting objects into the shopping basket, not replacing them on a shelf. We are augmenting Third Eye with a hand-tracking algorithm from the glasses camera to make the final picking action more robust, adding a text-in-the-wild recognition system to supplement the selected object's validation against the template image. We are also finessing the interface for shelf-depth estimation, which the user infers, so that the handoff from body movement to hand movement can be more seamless. These enhancements are the basis for developing a hybrid approach that integrates directional guidance for large-scale body movement based on the shelf view, small-scale body movement and arm and hand gestures based on a close-up view, and interfaces that provide a smooth transference from one view to the other.

Our combination of partial and complete system evaluations has given us enough insights to construct a clear path for refining Third Eye before testing with a larger number of PVIs. We have shown the current version at K–12 science fairs to inspire interest in computing applications, and conducted a hands-on demo for US Congress members in an effort to educate policymakers. We look forward to seeing results from a larger-scale test with Third Eye's enhanced version.

ACKNOWLEDGMENTS

This work is supported in part by NSF Expeditions: Visual Cortex on Silicon CCF 1317560.

REFERENCES

1. D. Dakopoulos and N.G. Bourbakis, "Wearable Obstacle Avoidance Electronic Travel Aids for Blind: A Survey," *IEEE Trans. Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 40, no. 1, 2010, pp. 25–35.
2. J. Nicholson, V. Kulyukin, and D. Coster, "ShopTalk: Independent Blind Shopping through Verbal Route Directions and Barcode Scans," *The Open Rehabilitation J.*, vol. 2, no. 1, 2009, pp. 11–23.
3. V. Kulyukin and A. Kutiyanawala, "Accessible Shopping Systems for Blind and Visually Impaired Individuals: Design Requirements and the State of the Art," *Open Rehabilitation J.*, vol. 3, 2010, pp. 158–168.
4. Y.T. Hsieh et al., "Designing a Willing-to-Use-in-Public Hand Gestural Interaction Technique for Smart Glasses," *Proc. ACM CHI Conf. Human Factors in Computing Systems* (CHI 16), 2016, pp. 4203–4215.
5. V. Lehtinen et al., "Dynamic Tactile Guidance for Visual Search Tasks," *Proc. 25th ACM Symp. User Interface Software and Technology* (UIST 12), 2012, pp. 445–452.
6. C. Siagian and L. Itti, "Rapid Biologically-Inspired Scene Classification using Features Shared with Visual Attention," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 2, 2007, pp. 300–312.
7. H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded Up Robust Features," *Proc. 9th European Conf. Computer Vision* (ECCV 06), 2006, pp. 404–417.
8. M. Merler, C. Galleguillos, and S. Belongie, "Recognizing Groceries in Situ Using in Vitro Training Data," *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (CVPR 07), 2007; www.researchgate.net/publication/221362362_Recognizing_Groceries_in_situ_Using_in_vitro_Training_Data.
9. R. Brenner, J. Priyadarshi, and L. Itti, "Perfect Accuracy with Human-in-the-Loop Object Detection," *Proc. European Conf. Computer Vision* (ECCV 16), 2016, pp. 360–374.
10. S. Brewster and L.M. Brown, "Tactons: Structured Tactile Messages for Nonvisual Information Display," *Proc. 5th Australasian User Interface Conf.* (AUIC 04), 2004, pp. 15–23.