

Improving Public Transit Accessibility for Blind Riders by Crowdsourcing Bus Stop Landmark Locations with Google Street View

Kotaro Hara¹, Shiri Azenkot², Megan Campbell², Cynthia L. Bennett², Vicki Le¹, Sean Pannella¹, Robert Moore¹, Kelly Minckler², Rochelle H. Ng², Jon E. Froehlich¹

¹Makeability Lab | HCIL
Department of Computer Science
University of Maryland, College Park
{kotaro, jonf}@cs.umd.edu

²DUB Group
Computer Science and Engineering
University of Washington, Seattle
{shiri, meganca}@cs.uw.edu

ABSTRACT

Low-vision and blind bus riders often rely on known physical landmarks to help locate and verify bus stop locations (*e.g.*, by searching for a shelter, bench, newspaper bin). However, there are currently few, if any, methods to determine this information *a priori* via computational tools or services. In this paper, we introduce and evaluate a new scalable method for collecting bus stop location and landmark descriptions by combining online crowdsourcing and Google Street View (GSV). We conduct and report on three studies in particular: (i) a formative interview study of 18 people with visual impairments to inform the design of our crowdsourcing tool; (ii) a comparative study examining differences between physical bus stop audit data and audits conducted virtually with GSV; and (iii) an online study of 153 crowd workers on Amazon Mechanical Turk to examine the feasibility of crowdsourcing bus stop audits using our custom tool with GSV. Our findings reemphasize the importance of landmarks in non-visual navigation, demonstrate that GSV is a viable bus stop audit dataset, and show that minimally trained crowd workers can find and identify bus stop landmarks with 82.5% accuracy across 150 bus stop locations (87.3% with simple quality control).

Categories and Subject Descriptors

H.5 [Information Interfaces and Presentation]: User Interfaces;
K.4.2 [Social Issues]: Assistive tech for persons with disabilities

General Terms

Measurement, Design, Experimentation, Human Factors

Keywords

Crowdsourcing accessibility; accessible bus stops; Google Street View; Mechanical Turk; low-vision and blind users

1. INTRODUCTION

For people who are blind or low-vision, public transportation is vital for independent travel [1,7,25,32]—particularly because their visual impairment often prevents driving. In previous formative work, we interviewed six blind adults about accessibility challenges in using public transportation [2]. We found that while buses were frequently a preferred mode of transit, determining the exact location of a bus stop was a major challenge [*ibid.*, p. 3249]. Strategies for finding bus stops included asking other pedestrians for information (if available) or locating known landmarks such as bus stop signs, shelters, or other physical objects (*e.g.*, benches).



Figure 1: Visually impaired travelers use landmarks to find and verify transit locations [2,14]. In this paper, we examine the feasibility of using Google Street View (GSV) and crowdsourcing to collect detailed information on bus stop locations and surrounding landmarks. The image above shows actual labels from crowdworkers in our Mechanical Turk study (Study 3). From left to right: blue circular icon=bus stop sign, magenta=bus stop shelter, yellow=bench, green=trash/recycling can.

In this paper, we focus specifically on the role of landmarks in helping blind and low-vision people find and identify bus stop locations. While some transit agencies provide brief descriptions of their bus stops online (*e.g.*, [26]), this information often lacks detail or is inaccessible to visually impaired riders—if available at all. Similar to our previous interview findings [2], the American Foundation for the Blind (AFB) notes that locating bus stops is a significant access barrier often because the bus stops are not clearly marked with non-visual indicators or are placed inconsistently off roadways [1]. The challenge of locating and identifying a bus stop is exacerbated when traveling to an unfamiliar location where both the bus stop placement and the position and type of surrounding landmarks are not known *a priori*.

To address this problem, we introduce and evaluate a new method for collecting bus stop location and landmark descriptions using online crowdsourcing and Google Street View (GSV). Using a custom tool that we built called *Bus Stop CSI (Crowdsourcing Streetview Inspections)*, crowd workers virtually navigate to and label bus stop signs and surrounding landmarks in GSV. This new approach is highly scalable in comparison to previous bus stop crowdsourcing work, *e.g.*, *GoBraille* [2] and *StopFinder* [29], which require users to describe bus stops *in situ* using a mobile device. While this paper focuses largely on data collection methods, we envision future work that integrates this data into transit agency websites and location-aware mobile transit tools such as *OneBusAway* [10]. For example, imagine a smartphone application that uses GPS and text-to-speech to automatically describe nearby and upcoming landmarks as a blind pedestrian navigates towards a bus stop.

We report on three studies beginning with an **interview study (Study 1)** of 18 people with visual impairments (7 with no functional vision) to inform the design of our crowdsourcing tool. These interviews extend and complement our aforementioned formative work [2] and further emphasize the importance of non-visual landmarks in helping blind/low-vision travelers find and verify a bus stop location. We then transition to describing two studies of GSV: a **comparative study (Study 2)** examining differences between physical bus stop audit data and audits conducted virtually with GSV, and an **online study (Study 3)** using Amazon Mechanical Turk (MTurk) designed to examine the feasibility of crowdsourcing bus stop audits using our tool.

In Study 2, we found a high correlation between our physical bus stop audit data and GSV images across four field sites in the Washington DC and Seattle metropolitan areas. This provides initial support for using GSV as a lightweight bus stop audit method. In Study 3, 153 MTurk crowd workers (turkers) labeled 150 bus stops using GSV via our custom tool. Overall, our results show that an individual turker is able to find and correctly label a bus stop and surrounding landmarks with 82.5% accuracy. This increases to 87.3% with simple 7-turker majority vote for quality control. While not perfect, these results point to the feasibility of using GSV and crowdsourcing to gather detailed bus stop descriptions. Future work should focus on training and quality control to increase accuracy.

In summary, the contributions of this paper are threefold involving both formative and summative findings: (i) our interview study adds to the existing literature on how blind and low-vision persons use bus transit, with a specific focus on navigating to and identifying bus stops; (ii) our comparative physical vs. virtual bus audit study is the first of its kind and establishes that GSV is a viable data source for collecting descriptions of bus stop features and surrounding landmarks; and, finally, (iii) our custom tool (*Bus Stop CSI*) and online crowdsourcing study shows that minimally trained crowd workers can find and describe bus stops using GSV with reasonable accuracy (> 82% without quality control).

2. RELATED WORK

Using public transit requires navigating a wealth of visual information from maps and schedules to bus stop markings and bus route signs. This reliance on visual information makes using public transit difficult for people with severe visual impairments [1,25]. With bus transit specifically, blind or low-vision persons can struggle with determining route and schedule information, purchasing fare, finding the correct bus stop location, getting on the appropriate bus, and getting off at the right stop [1,2,12,33].

Most transit tools designed to assist blind and low-vision bus riders focus on two issues: helping identify the correct bus to board when waiting at a bus stop [4,27] or providing alerts for an upcoming stop while riding the bus [21,22]. We are interested in addressing a prerequisite challenge: helping visually impaired riders find and verify bus stop locations through the use of physical landmarks and detailed bus stop descriptions (e.g., the presence of benches, bus shelters). In a survey of 55 persons with visual impairments, 85% reported difficulties in finding public transit pick-up points such as bus stops [12]. Recent work has emphasized the importance of physical landmarks in helping low-vision and blind users navigate to public transit [2,14]. Landmarks can only be used for navigation, however, when their location and spatial context (e.g., proximity to other physical objects) is known. Typically, this information is not captured or shared via traditional navigation tools (e.g., online maps).

Most relevant to our work is the *GoBraille* project [2] and its follow-up *StopFinder* [29], which was proposed but never evaluated. Both projects emphasized *in situ* mobile crowdsourcing to collect *and* present data about bus stops and surrounding landmarks to aid blind travelers (*ibid*, p. 323). Their *in situ* crowdsourcing approach takes advantage of the traveler's "downtime" while waiting for a bus: users fill out a simple form describing the bus stop (e.g., its location, relative direction, and encountered landmarks). While the reliance on blind users for bus stop data provides insights that are important to that community (e.g., non-visual perceptions of a landmark), the approach suffered from critical mass issues and data scarcity. While our aim is similar, our approach is unique: crowdsourcing data collection online using GSV where anyone at any time can contribute.

Omnidirectional streetscape imagery such as that found in GSV, Microsoft Bing Maps, and some Nokia Maps has been increasingly popular as a virtual audit technique in fields from urban informatics to public health research [3,8,15,17,18,30]. Reported benefits over physical audits include time-savings and the ability to monitor and analyze multiple cities from a central location [3,30]. As an emerging area of research, most work thus far has focused on examining agreement between virtual (e.g., GSV) and physical field audit data (e.g., [3,8,15,30]). Important for our work here, high levels of agreement have been found for measures including pedestrian safety, traffic and parking, and pedestrian infrastructure. To our knowledge, however, no one has specifically looked at the concordance between physical and virtual audit data for bus stops and their surrounding environment (which is the focus of Study 2).

With regard to crowdsourcing for accessibility, Bigham and colleagues argue that current technological infrastructure provides unprecedented access to large sources of human power that can be harnessed to address accessibility challenges [6]. Recent examples of such crowdsourcing systems include VisWiz [5] and Legion:Scribe [24]. More relevant to our work is Tiramisu [31], a mobile crowdsourcing tool developed via universal design to help gather and disseminate information about bus arrival time and capacity. Our approach is complementary but does not rely on mobile crowdsourcing or continuous, active use by crowd workers to provide benefits. Finally, in the last decade, a growing number of crowdsourcing systems dedicated to geographic content have emerged (e.g., Wikimapia, OpenStreetMap, and Cyclopath [28]). Interestingly, past work has found that user-contributed map data quality is high even when compared to proprietary systems (e.g., [11,16]). Though we currently rely on paid labor via MTurk, we plan to explore community-sourcing and volunteer contributions.

3. STUDY 1: FORMATIVE INTERVIEWS

In 2010¹, we conducted formative interviews with six blind adults to learn about the challenges faced by visually impaired persons when using public transit [2]. Here, we extend upon and complement this previous work by covering a wider variety of transit systems and involving a greater diversity of visually impaired participants. In addition, we specifically investigate the role of non-visual landmarks in bus stop navigation.

3.1 Interview and Analysis Method

We recruited 18 participants (10 male) with visual impairments from the US and Canada with an average age of 52.1 ($SD=12.0$; $range=24-67$). Eleven participants could not easily read street signs due to their visual impairment. Of these, 7 had no functional vision. As bus transit systems differ across population densities,

¹ Interviews were done in 2010, but the findings were published in 2011.

we sought participants from different neighborhood types: 8 participants lived in urban areas, 7 suburban, and 3 in small towns. Participants were recruited via mailing lists affiliated with blindness organizations and were paid \$15. The recruitment email explicitly stated that we were investigating public transit accessibility and that participants must be blind or low-vision.

We conducted semi-structured, phone-based interviews, which lasted ~40 minutes. We asked participants about patterns of public transit use, challenges experienced therein, and coping/mitigation strategies. We then described a hypothetical smartphone application that provided the location and description of bus stops and surrounding landmarks (e.g., via GPS tracking and text-to-speech). We asked participants to assess the importance of various landmarks for this software application. We recorded, transcribed, and coded the interviews using an open coding methodology. While our interviews covered a broad range of subjects related to transit accessibility, below we primarily concentrate on findings related to locating bus stops.

3.2 Bus Stop Related Interview Findings

For most participants, public transit was critical for daily mobility. One woman, for example, stated that the lack of accessible public transit “*played into her decision*” to retire. Similar to prior work [2,12,33], participants described many challenges when using public transit including finding bus stops, knowing which bus to board, and when to disembark.

Most relevant to this paper, half of the participants experienced difficulty finding the exact location of bus stops when travelling. Difficulties included determining the specific location of a bus stop (e.g., near-side of intersection, half-way down the block), obtaining accessible information sources, and knowing which landmarks and businesses indicate a proximal bus stop. Because bus stop designs and placement can vary widely within a city—from stops with a myriad of physical landmarks (e.g., shelters, benches, trash cans, and newspaper boxes) to stops with only a pole—one participant said with frustration:

There's really no rhyme or reason of where they put bus stops. And there's no way to...tell where a bus stop [is], 'cause you don't ever know where the pole is, or how it's marked, or... anything like that. (P3, age=63, blind)

For this participant, the main reason he did not use public transit was because of the challenges he faced in finding bus stops. Another participant noted that some stops in his city were hard to find because they had no non-visual landmarks, only painted curbs. Many noted that consistent stop locations and landmarks would significantly help them overcome this accessibility challenge. For both blind and low-vision participants, finding an unfamiliar stop took a lot of time and, as one participant explained, required adjusting expectations to reduce stress:

I think also just not to worry about it so much. Just not stress out about it. Just know that it will be new and it will take a little more time to figure it out. (P14, 55, blind)

To find bus stops, participants mentioned using walking directions from transit trip planners (if available in an accessible form), calling the transit agency², or asking a sighted person questions about the stop's location. Ten participants (53%) reported asking pedestrians or other transit riders for information—a strategy only available when others are present (i.e., more difficult at night or in more rural areas). Some participants used orientation or mobility

instructors to help guide them to routine bus stops. Once participants reached the vicinity of the stop, they commonly searched for landmarks. For example, if a person uses a cane, s/he can hear an echo from a shelter when walking by.

When asked about which landmarks at bus stops are most important to navigation, participants identified shelters and benches as the most helpful followed by trash cans, newspaper bins, grass shoulders, and other non-visual indicators. A few also mentioned knowing the *shape* of the bus stop pole (e.g., thin vs. thick, two-column vs. one). One participant emphatically stated that all landmark information would be of critical importance. Five participants also mentioned the importance of knowing nearby businesses because of their distinct sounds and smells.

I look for landmarks... like a bus shelter at a certain place... or if there's a hedge, like bushes in front of a certain place and right by those bushes there's a newspaper rack or something like that then I know that it's my stop. If it's in front of a coffee shop...if there's a hotdog stand there, then I know that the bus stop is in front of the hot dog stand, you smell it... Noises too, you know different sounds. (P14, 55, blind)

Though participants relied on various technologies for planning a trip on public transit, only five participants (26%) used smartphone applications for such tasks. These applications provided either real-time or scheduled arrival information, and helped participants determine which bus to board. None of our participants used technology tools while walking around and looking for stops, perhaps because no such tool yet exists.

3.3 Study 1 Summary

In summary, although our first interview study was conducted three years ago [2], the major challenges of blind and low-vision public transit riders remain the same despite technological improvements in navigation tools, smartphone applications, and accessible bus systems (e.g., automated announcements). Most participants said that having information about landmarks would enable them to use transit more easily (even five participants who could sometimes read street signs). Descriptions of the shape and location of bus stop poles, shelters, and benches as well as information indicating their presence seemed most beneficial.

4. STUDY 2: PHYSICAL vs. GSV AUDITS

To assess the viability of using GSV to audit bus stops, we needed to first establish that the bus stops captured in the GSV image dataset do not differ significantly from current reality (e.g., because of image age). Thus, in Study 2, we conducted both in-person bus stop audits and GSV-based audits across the same four target geographic areas and compared the results. An audit here means logging the existence of landmarks at bus stops using a predefined codebook (described in Section 4.2). While the primary aim of this study was to explore what differences, if any, would exist between the GSV and physical bus stop audit data, we had two secondary aims. First, to investigate the feasibility and difficulty of the audit task itself (e.g., can members of our research team agree amongst themselves on the application of audit measures across various bus stop scenes?). Second, to produce a high-quality *ground truth* dataset that could be used to assess crowd worker performance in Study 3.

Our bus stop audit sites included four neighborhoods in the Washington DC and Seattle, WA metropolitan areas (Table 1; Figure 2). As bus stop designs differ across cities and neighborhoods, we selected a range of densities (e.g., downtown vs. suburban) and neighborhood types (e.g., residential vs. commercial). Additionally, we emphasized areas that have high-demand for public transit (e.g., including schools, major

² In our prior work, one participant noted poor experience with calling transit agencies because they could not adequately explain bus stop locations over the phone (perhaps because the agency itself did not store sufficiently detailed descriptions about their bus stops in their database) [2].

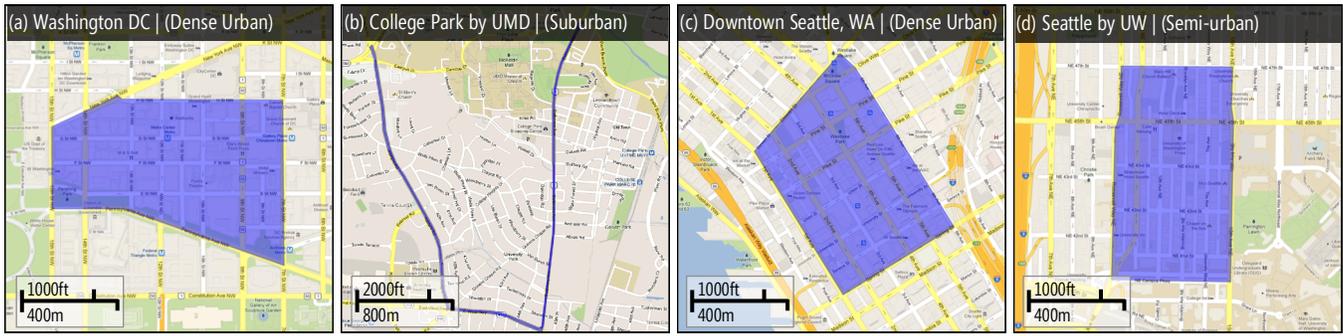


Figure 2: The four audit areas used in Study 2 and 3 spanning a range of neighborhood types in Washington DC and Seattle, WA. In all, we surveyed 179 bus stops across 42.2 linear km. Each field site took approximately one day to physically survey (~6 hours).

department stores, convention centers, and museums). These same areas are also used in our crowdsourcing audit study (Study 3).

4.1 Collecting Physical Audit Data

Two separate research teams physically visited the bus stop locations: one team in Seattle and the other in Washington DC. Teams walked (or biked) down each street in the predefined study area. They carried smartphones with GPS to help navigate to and track bus stops. An online spreadsheet prefilled with bus stop locations (*e.g.*, Baltimore & Campus Dr.) and a Google Map URL allowed the researchers to track their position and the target bus stop in real-time on an interactive map. Visited stops were marked in the spreadsheet and linked to a unique index for later analysis.

At each bus stop location, we took 7-10 geo-timestamped pictures from varying angles—roughly 360° around the bus stop from the sidewalk and street (far more angles than GSV)—and analyzed them *post hoc*. We were careful to capture clear images without occlusion problems. This photographic approach had two primary advantages: it created an image dataset analogous to GSV, which allowed us to apply a similar auditing methodology to both, and it allowed us to examine the image dataset multiple times without returning to the field site.

4.2 Auditing Methodology

While we had two separate teams photograph bus stops during the in-person field site visits, we used one single team of three researchers to independently audit (code) both the physical and GSV image datasets. This reduced confounds due to different auditors. Although bus stop auditing may seem like an objective process, it is, in fact, *subjective* and requires following a qualitative coding methodology. For example, one auditor may simply miss seeing a particular object in a scene or may misperceive or mislabel one object as another. By following the iterative coding method from Hurschka *et al.* [19], our aim was to produce two high-quality audit datasets—one for each image dataset: physical and GSV—that could then be compared.

To begin the auditing process, an initial codebook was derived for each bus stop landmark: (i) bus stop signs, (ii) bus stop shelters, (iii) benches, (iv) trash/recycling cans, (v) mailbox and newspaper bins, and (vi) traffic signs and other poles. These landmarks were selected based on the findings from our interviews as well as from bus stop design guidelines (*e.g.*, [9]). The codebook provided detailed definitions of each along with visual examples. We also defined the audit area around a bus stop as 20 feet (~6.1 meters) in either direction from the bus stop sign (from [20]). Note, however, that as our audits were performed via visual inspection of images (for both the physical and GSV datasets), auditors could only estimate distances.

	Washington DC	College Prk by UMD	Downtown Seattle	Seattle by UW	Overall
Description of Audit Area	Dense urban	Suburban (next to U. of Maryland)	Dense urban	Semi-urban (next to U. of Washington)	N/A
Total Linear km Surveyed	11.2	11.9*	8.0	11.1	42.2
# of Bus Stops Found in Physical Audit	82	36	35	26	179
# of Bus Stops Found in Physical Audit but Missing from Google Maps API	21	4	3	1	29
Avg. GSV Data Age (SD)	1.9 yrs (0.3)	1.0 yrs (0.7)	1.9 yrs (0.3)	2.1 yrs (1.1)	1.75 (0.7)

Table 1: The four areas surveyed in our physical and virtual (GSV) audits. Total Linear km Surveyed represents unidirectional surveying distance except for the *, which is bidirectional because of wider streets separated by a median (*i.e.*, auditors walked/biked one side of road and then the other). See Figure 2.

During auditing, count data was entered into a preformatted spreadsheet tracking the number of each landmark at each bus stop. As prescribed by Hruschka *et al.*, each auditor began by independently coding a small subset of data—in this case, 15 DC and 5 College Park bus stop locations. Afterwards, the auditors came together to discuss and modify problematic codes. With the updated codebook, the entire DC and College Park physical image dataset was audited (including the original 20 locations, which were re-audited) followed by the GSV dataset. We conducted a similar iterative coding process for the two Seattle audit areas. The codebook descriptions were updated to reflect Seattle bus stop designs.

The GSV audits differed from the physical image dataset audits in two ways: first, the auditors used a GSV interface where they could control camera angle and location rather than browse through a set of static images; second, the auditors rated the overall difficulty of auditing each location on a 5-point Likert scale, where 1=*very easy to assess* and 5=*very hard to assess*. These ratings will be used later in Study 3 to investigate whether crowdsourcing audit accuracy changes based on rated difficulty.

4.3 Inter-Rater Agreement on Audit Data

Before comparing the physical audit data to the GSV audit data (Section 4.4), we needed to first calculate inter-rater agreement *between researchers* for each individual dataset. For this, we applied the Krippendorff’s Alpha (α) statistical measure (see [23]). Although we have previously used Fleiss’ kappa to compute inter-rater agreement on streetscape audit tasks [18], this statistical measure cannot be applied to count data, which is what we have here. Our results are presented in Table 3 (1st pass columns). The overall α score between researchers was 0.909 for the physical audit dataset and 0.850 for the GSV audit dataset.

Similar to most statistical measures for inter-rater agreement, there is no universally accepted threshold for determining high agreement with the Krippendorff’s Alpha measure. However, [23] suggests that agreement scores of $\alpha \geq 0.800$ are generally considered reliable while data below $\alpha < 0.667$ should be discarded or recoded (p. 241). Though none of our α scores fell below 0.667 for either dataset, some categories had $\alpha < 0.800$. One primary source of disagreement involved differing perceptions of what geographic area constituted a bus stop (recall the 20ft perimeter). For some bus stop locations, traffic signs, poles, and other landmarks extended just beyond or just within the prescribed bus stop range. These edge cases were difficult to assess and contributed to the lower α score. Note also that the GSV agreement scores were lower on average than the physical audit dataset often because of inferior-quality images (e.g., the GSV privacy protection algorithm misidentified some bus stop signs as vehicle license plates and blurred them out; see [13]).

To alleviate such disagreements as recommended by Hruschka *et al.* [19], the three auditors discussed low agreement codes (any $\alpha < 0.800$) and updated the codebook once again. The auditors then took a 2nd full independent pass both on the physical and GSV audit datasets but focused only on those bus stop landmarks that previously had an α score < 0.800 . The updated results are in Table 2 (2nd pass columns). On this 2nd pass, the overall agreement increased from 0.909 to 0.944 for the physical audit dataset and 0.850 to 0.930 for the GSV dataset. Importantly, all α scores were now ≥ 0.800 thereby completing our iterative coding scheme.

Bus Stop Landmark	Physical Audit Image Dataset		GSV Audit Image Dataset	
	1 st Pass (α)	2 nd Pass (α)	1 st Pass (α)	2 nd Pass (α)
Bus Stop Sign	0.937	0.972	0.761	0.916
Bus Stop Shelter*	0.991	0.991	0.955	0.955
Bench*	0.940	0.940	0.870	0.870
Trash/Recycling Can	0.876	0.886	0.793	0.946
Mailbox/ Newspaper Bin	0.886	0.957	0.768	0.938
Traffic Signs/ Other Poles	0.683	0.866	0.685	0.874
Overall	0.909	0.944	0.850	0.930

Table 2: Krippendorff’s Alpha inter-rater agreement scores between three researchers on both the physical audit and GSV audit image datasets. Following Hruschka *et al.*’s iterative coding methodology [19], a 2nd audit pass was conducted with an updated codebook for low-agreement scores—in our case, $\alpha < 0.800$. *Excluded categories for 2nd coding pass because original agreement $\alpha \geq 0.800$.

4.4 Comparing Physical vs. GSV Audit Data

The high agreement scores within the physical and GSV datasets provides evidence that the audit data is consistent and of good quality. Consequently, we can move towards examining the key research question of Study 2: how does the physical audit dataset compare to the GSV dataset? To investigate this question, two more small procedural steps are required: first, we need to amalgamate the three-auditor count data into a single count set for both datasets and then, second, we need to decide upon some mathematical approach to compare them. For the amalgamation method, we take the median of the three auditor counts for each bus stop landmark at each bus stop location. For example, if R1 found 1 traffic sign at a specific bus stop location, R2 found 4 traffic signs, and R3 found 5, then the median count between them would be 4. This allowed us to transform the three count datasets into one for both the physical and GSV audit data. For the comparison method, similar to Rundle *et al.* [30], we calculate a Spearman rank correlation between the two count sets (physical and GSV).

Our results are presented in Table 3; all are statistically significant at $p < 0.001$. Using Rundle *et al.*’s definition of high correlation, all of our landmark coefficients (ρ) are highly correlated ($\rho > 0.60$) between the physical and GSV datasets. The two highest are for bus stop infrastructure: *Bus Stop Shelters* ($\rho=0.88$) and *Benches* ($\rho=0.88$). The two lowest are *Bus Stop Signs* ($\rho=0.61$), which are sometimes difficult to see in GSV, and *Trash/Recycling Cans* ($\rho=0.72$), which are likely to be the most transient landmark type (e.g., they may move a lot over time).

Physical vs. GSV Audit Data	Bus Stop Sign	Bus Stop Shelter	Bench	Trash / Recycling	Mailbox / News. Bins	T. Signs / Other Poles
Coefficient (ρ)	0.612	0.877	0.875	0.715	0.776	0.811

Table 3: Following Rundle *et al.* [30], we performed a Spearman rank correlation between the physical and GSV bus stop landmark count audit datasets. For all coefficients (ρ), $p < 0.001$.

It is important to note that during the physical audit, we encountered 29 bus stops that were *not* in Google Transit’s bus stop location dataset (21 of which were in downtown Washington DC); see Table 1. This Google transit dataset is independent of the GSV images. Only *three* of these bus stops, however, were also missing in GSV (due to outdated images). The above correlation results are for all 179 physical audit locations with zeros filled in for the three missing bus stops in the GSV datasets.

4.5 Study 2 Summary

In summary, Study 2 demonstrates that bus stop auditing is a subjective process but, more importantly, that the GSV audit dataset is highly correlated with the physical audit dataset. This indicates that despite instances of GSV image ages being over two years old, GSV is a viable data source for gathering up-to-date information on bus stop locations and surrounding landmarks.

5. OUR BUS STOP LABELING TOOL

Shifting now to preparations for our third study: to allow crowd workers to examine and describe bus stops and surrounding landmarks in GSV, we created an interactive online labeling tool called *Bus Stop CSI (Crowdsourcing Streetview Inspections)* in Javascript, PHP, and MySQL. Unlike previous crowdsourcing GSV work which uses static imagery to collect labels (e.g., [15,17,18]), our labeling interface is fully interactive and allows the crowd worker to move about and control the camera view in the 360 degree GSV panoramic space (see Figure 3). Although this interactive freedom increases task complexity, the benefits are twofold: first, the crowd worker can “walk” in GSV to find the target bus stop; second, the crowd worker can shift their view to find an optimal labeling perspective (e.g., a camera view that avoids occlusions). As we deployed our tool on MTurk, the description below is written for that context.

5.1 Using the Bus Stop Labeling Tool

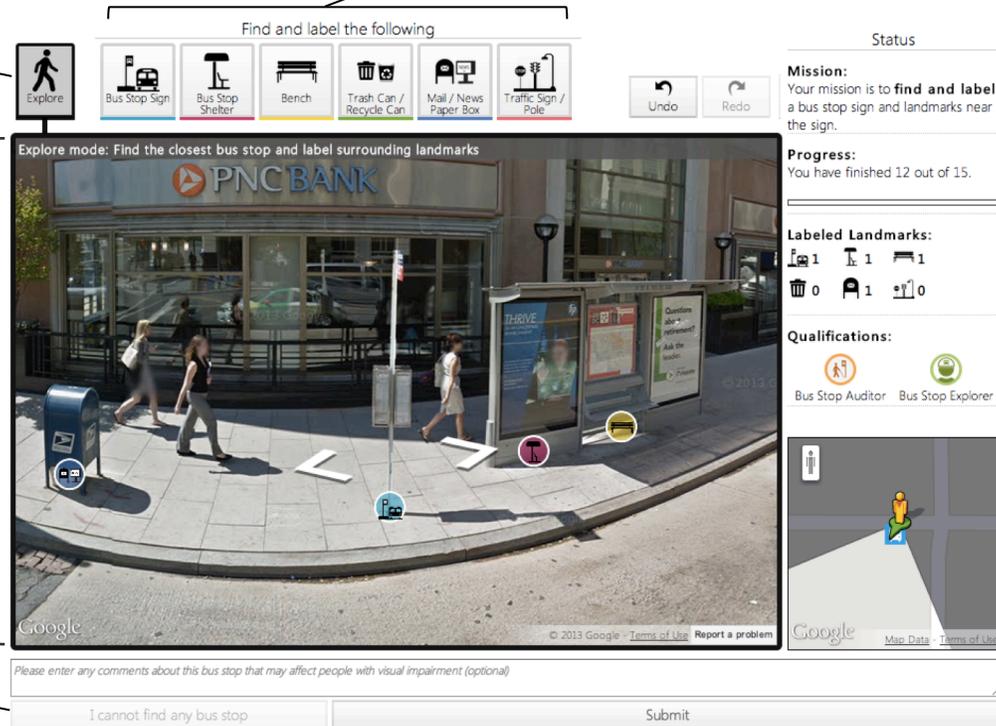
When a turker accepts our HIT, they are immediately greeted by a four-stage interactive tutorial. Each stage is dedicated to progressively teaching the turker about some new interaction or labeling feature in our tool (e.g., how to pan the camera, walk, and label). The tutorials also teach higher-level concepts such as how sometimes a target bus stop icon may show up in the 2D-map view but may not actually exist in the GSV pane (due to inaccuracies in Google’s transit data). Turkers must successfully complete one stage before moving on to the next. Because the bus stop signs and landmarks differ in look and feel across cities, we created separate interactive tutorials for Washington DC and Seattle (eight interactive in total; four for each metropolitan area). If a turker was trained in one city, they were required to retrain in the other city.

When one of these “label” buttons is selected, the interface enters the **Labeling Mode**. The mouse cursor turns into a representative icon for the selected label type. The user directly clicks on the object in the GSV pane below to place the label. In this mode, unlike the Explore Mode, the camera angle and location is fixed. The interface automatically returns to Explore Mode after each label is placed.

The **Explore Mode** (currently selected) allows the user to control the GSV camera angle and “walk” up to two steps in any direction beyond the drop point.

The **GSV pane** is the primary interaction area for exploring and labeling.

If the user cannot find a bus stop in the scene, they can click this button and provide details.



The **Status side panel** provides details on the user’s progress and their qualification badges (which they earn in our interactive tutorials).

The user’s location and view direction are represented in this **top-down 2D map view**. The bus stop icon (in blue) is drawn based on location data from Google.

The user clicks the **Submit button** to upload their labels.

Figure 3: The Bus Stop CSI interface. We use the Google Maps Transit API to determine drop locations nearby bus stops. Crowd workers use the *Explorer Mode* to move around and look for the target bus stop (indicated by the blue icon in the 2D-map view) and the *Labeling Mode* to label any of the six bus stop landmark types. Clicking the Submit button uploads the labels (in this case, a mailbox, bus stop sign, shelter, and bench). The worker is then transported to a new location unless the HIT is complete (14-16 bus stop locations are included in each HIT).

Once the tutorials are successfully completed, we query the Google Maps API to automatically drop the turker close to a bus stop in the audit area and the task begins in earnest. Bus Stop CSI has two primary modes of interaction: the *Explorer Mode* and the *Labeling Mode*. In the Explorer Mode, the user interacts in the GSV pane using the traditional Street View inputs. Walking is controlled by clicking the arrow movement widgets (<, >, v, and ^). Horizontal and vertical panning in the 360 degree view is controlled by clicking and dragging the mouse across the image. When the user is first dropped into a scene, s/he is defaulted into Explorer Mode. When the user clicks on one of the six labeling buttons, the interface switches automatically to the Labeling Mode. Here, mouse interactions no longer control movement and camera view. Instead, the cursor changes to represent the currently selected label. The user can then apply the selected label by clicking on the appropriate landmark in the GSV pane. Our tool automatically tracks the camera angle and repositions the applied labels in their correct location as the view changes—in this way, the labels appear to “stick” to their associated landmark. Turkers cannot see previously entered labels by other workers.

In early pilot studies, we found that users would get disoriented by accidentally “looking” straight down (towards the street) or straight up (towards the sky) in the GSV pane. Thus, to simplify GSV interaction and to focus the view appropriately on street-level features, we reduced vertical panning to 20 degrees (0, -20). Other GSV adjustments include: hiding the onscreen camera control and zooming widgets, disabling keyboard interactions (to prevent accidental movement), and hiding textual overlays (e.g., street names). In addition, we prevented users from moving more

than two steps in any direction away from their initial drop point. This constraint prevented users from unnecessarily walking down streets in search of bus stops. In our dataset, a single GSV “step” translates to roughly 5-10 meters of real-world movement (GSV steps are smaller in denser areas).

6. STUDY 3: CROWDSOURCING LABELS

To investigate the potential of using minimally trained crowd workers to find and label bus stop landmarks, we posted our tool to MTurk in April 2013. In each HIT, turkers needed to label 14-16 bus stop locations. We paid \$0.75 per HIT (\$0.047-0.054 per labeling task); which was decided based on the task completion time in pilot studies (e.g., approximately \$0.10 per minute). Although we used 179 bus stop locations in Study 2, here, we use a subset 150. This subset is necessary because, as previously mentioned, 29 bus stop locations do not show up in the Google Maps Transit API (see Table 1). We use this API to automatically place turkers next to bus stops in our labeling tool. If the API is unaware of the bus stop, we cannot determine its location.

6.1 Assessing Accuracy

In order to assess turker performance, we need ground truth data about which landmarks exist at each bus stop location. For this, we use the median count GSV dataset from Section 4.4. Recall that to produce this consolidated dataset, we calculated the median count of each landmark type from the three auditor datasets across every bus stop location. Here, we further transform these counts into binary presence indicators for each landmark type. In other words, our ground truth dataset is a 150 row (for bus stop locations) x 6 column (for landmark types) matrix where $cells=1$ represent the presence of that landmark type at the specified bus

stop and $cells=0$ represent an absence. Although the *Bus Stop CSI* tool gathers raw landmark counts and relative location data on landmarks (e.g., a trashcan is north of the bus stop sign), we do not evaluate this level of granularity here. Thus, our analysis focuses only on whether crowd workers properly indicated the presence/absence of a landmark in a scene but without regard for multiple occurrences. We leave more sophisticated assessments for future work.

6.2 High Level Results

In total, 153 distinct turkers completed 226 HITS (3,534 labeling tasks) and provided 11,130 bus stop landmark labels. On average, turkers completed 1.48 HITS ($SD=1.17$), which is equivalent to 23.1 labeling tasks ($SD=19.0$). The median labeling time per task was 44.7s ($avg=71.8s$; $SD=213.1s$) and the average number of labels per panoramic image was 3.15 ($SD=3.06$). When compared with our ground truth dataset, overall turker accuracy was 82.5% ($SD=0.3\%$) for properly detecting the presence/absence of a landmark across the 150 bus stop locations.

When broken down by landmark type (Table 4), the mailbox/newspaper bin landmark type followed by the bus stop shelter and bench had the highest accuracies at 88.8% ($SE=0.4\%$), 88.6% ($SE=0.5\%$), and 83.3% ($SE=0.5\%$) respectively. These all tend to be fairly salient landmark types in GSV. In contrast, the lowest scoring landmark type (Traffic Signs / Other Poles at 66.2%) is the most open-ended label (i.e., least defined) making it susceptible to confusion and misuse. This is particularly true given that our ground truth data had a constrained 20 foot extent on either side of the bus stop sign meaning that potentially correct labels placed beyond that area could be flagged as incorrect. In the future, we plan to account for distance in our assessments.

	Bus Stop Sign	Bus Stop Shelter	Bench	Trash/ Recycling	Mailbox/ Newspaper Bins	T. Signs / O. Poles	Overall
Avg. Accuracy	81.9%	88.6%	83.3%	84.9%	88.8%	66.2%	82.5%
(N=153 turkers)	(0.6)	(0.5)	(0.5)	(0.4)	(0.4)	(0.4)	(0.3)

Table 4: The average labeling accuracy with one turker per scene across all 150 bus stops. Cell format: Average (Standard Error).

Returning to the researcher supplied difficulty ratings from Study 2, we found a significant difference ($p < 0.0001$) between turker performance on bus stop locations rated easy by our research team ($N=116$) vs. those rated medium-to-hard ($N=34$). For the easy locations, our average per turker accuracies were 84.5% ($SE=0.3\%$). This decreased to 74.3% ($SE=0.7\%$) for the hard locations, which suffered from occlusion, blurred images, and required more movement (including a scene where one virtual step leapt forward in a disorienting manner).

6.3 Accuracy as a Function of Majority Vote Size

Collecting multiple labels per bus stop location helps account for the natural variability of human performance and reduces the influence of occasional errors; however, it also requires more workers. Similar to Hara *et al.* [18], here we explore accuracy as a function of turkers per scene. We recruited 21 (or more) turkers for each of the 150 bus stop locations. We compare ground truth data with majority vote labels across four turker groups: 1, 3, 5, and 7. Because we have at least 21 turkers per bus stop location, we can compute accuracies multiple times for each group size, average the results, and calculate error margins. The overall goal, here, is to produce a more accurate portrayal of expected future performance for each group size. For example, when we set the majority vote group size to three, we randomly permute seven groups of three turkers. In each group, we calculate the majority vote answer for a given bus stop location in the dataset and compare it with ground truth. This process is repeated across all

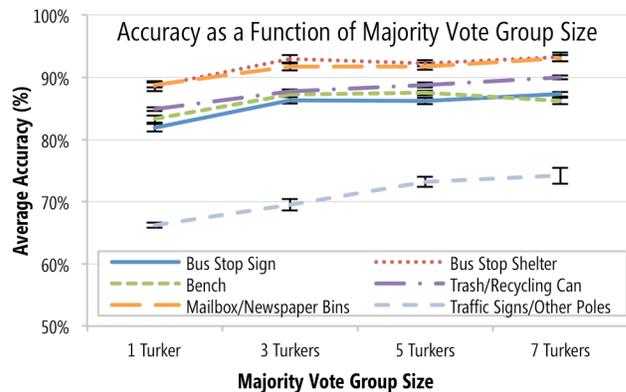


Figure 4: Overall accuracy as a function of majority vote group size. Each graph point is based on multiple permutations of the majority vote group size across all 150 bus stop locations. Standard error bars are in black. Note: the y-axis does not start at 0% (range: 50-100%).

locations and the five group sizes, where (X =majority vote group size, Y =number of groups): (1,21), (3, 7), (5,4), (7, 3). See [18].

Overall, we found that accuracy does indeed increase with majority vote group size from 82.5% to 85.8% with 3 turkers and 87.3% with 7 turkers. These gains, in general, diminish in magnitude as majority vote group size grows (Figure 4). However, for the hardest landmark label type (*Traffic Signs / Other Poles*), we see a continued steady increase as the majority vote size grows—perhaps indicating wisdom in the crowds for more challenging landmark types.

6.4 Study 3 Summary

In summary, though our current crowdsourcing experiments and analyses are rather simple, they are the first results to demonstrate that minimally trained crowd workers can accurately find and label bus stop landmarks in GSV ($> 82\%$). Future work should focus on more sophisticated analyses of worker labels including count and placement accuracy in each scene. In addition, more work is needed to establish the required accuracy level needed to provide value to transit agencies and navigation tools.

7. DISCUSSION AND CONCLUSION

In this paper, we conducted three studies related to the use of non-visual landmarks in locating and verifying bus stops. While Study 1 extended upon our previous formative work, our findings re-emphasized the significance of landmarks in aiding visually impaired navigation. For example, we found that benches and shelters were most helpful, which crowd workers correctly labeled 83.3% and 88.6% of the time, respectively, in Study 3—such a result demonstrates the interconnections between our studies. Study 2 showed that despite data age and occlusion problems, GSV could be used as a lightweight dataset for bus stop audits (even when compared to physical audit data). Finally, and perhaps most importantly, Study 3 showed that a minimally trained crowd worker could find and label bus stops in *Bus Stop CSI* with 82.5% accuracy, which jumps to 87.3% with a simple 7-turker majority vote scheme). Taken together, these three studies advance the current literature and understanding of how information about bus stop landmarks could be potentially collected and used to guide low-vision and blind bus riders. With that said, our work is not without limitations. Here, we briefly discuss limitations that could affect the scalability and accuracy of our approach.

Inaccurate bus stop locations. While our physical audit in Study 2 found 179 bus stops, 29 of these were missing from the Google Maps API. Because we rely on this same API in our *Bus Stop CSI*

tool, these 29 bus stops could not be visited—even if they were visible in GSV (in this case, all but three were). Similarly, often times we found that the exact location of bus stops in the Google Maps API was inaccurate (e.g., wrong place on the block, wrong side of an intersection). This made our 2D-map pane confusing for some scenes—a worker would point the avatar toward the bus stop icon but would not see a bus stop in the GSV pane. Other data sources (e.g., OpenStreetMap) could likely be used to mitigate this problem.

Image age. While we observed high concordance between our GSV bus stop audit data and our physical audit data, the image age in GSV remains a concern. Although Google does not publicly specify a GSV update plan from city-to-city, Washington DC has been updated three times in the last four years. In addition, Google just updated 250,000 miles of road in early October 2012 (<http://goo.gl/hMnM1>).

Scene difficulty: The following GSV-related problems made it more challenging to label bus stops: (i) *distance*: most streets are driven once by a GSV car from a single car lane in one direction. This can create distant views of bus stops; (ii) *occlusion*: bus stop landmarks are sometimes occluded by a parked bus or other obstacle, (iii) *lighting*: shadows from trees and buildings can make bus stop landmarks hard to see; and (iv) *blur*: as previously mentioned, sometimes GSV misidentifies a bus stop sign as a license plate and blurs it out, which makes it harder to identify. One potential solution would be to integrate other streetscape imagery sources (e.g., Microsoft Streetside) to gain multiple simultaneous views of an area.

Selecting bus stop landmarks. Our tool allowed crowd workers to label six landmark types but other landmark types could also be useful (e.g., grass, trees). For example, one turker left a comment saying, “There is a tree very close to the bus stop sign” Future work should examine other landmark types and continue performing user-centered design to see how these landmarks affect navigation.

8. ACKNOWLEDGEMENTS

This work was supported by an NSF grant (IIS-1302338) and a Google Faculty Research Award.

9. REFERENCES

1. American Foundation for the Blind. Accessible Mass Transit. 2013. <http://www.afb.org/section.aspx?SectionID=40&TopicID=168&DocumentID=907>.
2. Azenkot, S., Prasain, S., Borning, A., Fortuna, E., Ladner, R.E., and Wobbrock, J.O. Enhancing independence and safety for blind and deaf-blind public transit riders. *Proc of CHI'11*, 3247–3256.
3. Badland, H.M., Opat, S., Witten, K., Kearns, R.A., and Mavoa, S. Can virtual streetscape audits reliably replace physical streetscape audits? *Journal of Urban Health*, 87, 6 (2010), 1007–16.
4. Banâtre, M., Couderc, P., Pauty, J., and Becus, M. Ubibus: Ubiquitous Computing to Help Blind People in Public Transport. *Proc of MobileHCI'04*, 310–314.
5. Bigham, J.P., Jayant, C., Ji, H., et al. VizWiz: nearly real-time answers to visual questions. *Proc of UIST'10*, 333–342.
6. Bigham, J.P., Ladner, R.E., and Borodin, Y. The design of human-powered access technology. *Proc of ASSETS'11*, 3–10.
7. Blind Citizens Australia. BCA Public Transport Policy. 1994. <http://goo.gl/lkSkw>
8. Clarke, P., Ailshire, J., Melendez, R., Bader, M., and Morenoff, J. Using Google Earth to conduct a neighborhood audit: reliability of a virtual audit instrument. *Health & place* 16, 6 (2010), 1224–9.

9. Easter Seals. *Toolkit for the Assessment of Bus Stop Accessibility and Safety*. 2008.
10. Ferris, B., Watkins, K., and Borning, A. OneBusAway: results from providing real-time arrival information for public transit. *Proc of CHI'10*, 1807–1816.
11. Flanagan, A.J. and Metzger, M.J. The credibility of volunteered geographic information. *GeoJournal* 72, 3-4 (2008), 137–148.
12. Golledge, R.G., Marston, J.R., and Costanzo, C.M. Attitudes of Visually Impaired Persons Toward the Use of Public Transportation. *Journal of Visual Impairment and Blindness* 91, 5 (1997), 446–459.
13. Google. Google Street View Privacy. 2013. <http://goo.gl/UoBNp>.
14. Guentert, M. Improving public transit accessibility for blind riders: a train station navigation assistant. *ASSETS'11 Poster*, 317–318.
15. Guy, R. and Truong, K. CrossingGuard: exploring information content in navigation aids for visually impaired pedestrians. *Proc of CHI'12*, 405–414.
16. Haklay, M. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Env. & Planning B: Planning & Design* 37, 4 (2010), 682–703.
17. Hara, K., Le, V., and Froehlich, J. A feasibility study of crowdsourcing and google street view to determine sidewalk accessibility. *ASSETS'12 Poster*, 273–274.
18. Hara, K., Le, V., and Froehlich, J. Combining Crowdsourcing and Google Street View to Identify Street-level Accessibility Problems. *Proc of CHI'13*.
19. Hruschka, D.J., Schwartz, D., St.John, D.C., et al. Reliability in Coding Open-Ended Data: Lessons Learned from HIV Behavioral Research. *Field Methods* 16, 3 (2004), 307–331.
20. Intercity Transit. *Bus Stop Specification Guidelines*.
21. Jacob, R., Shalaih, B., Winstanley, A., and Mooney, P. Haptic Feedback for Passengers Using Public Transport. In H. Cherifi, J. Zain and E. El-Qawasmeh, eds., *Digital Information and Comm Tech and Its Applications*. Springer Berlin Heidelberg, 2011, 24–32.
22. Kostianen, J., Erkut, C., and Piella, F.B. Design of an audio-based mobile journey planner application. *Proc of MindTrek'11*, 107–113.
23. Krippendorff, K.H. *Content Analysis: An Introduction to Its Methodology*. Sage Publications, Inc, 2003.
24. Lasecki, W., Miller, C., Sadilek, A., et al. Real-time captioning by groups of non-experts. *Proc of UIST'12*, 23–34.
25. Marston, J.R. and Golledge, R.G. The Hidden Demand for Participation in Activities and Travel by Persons Who Are Visually Impaired. *J. of Visual Impairment & Blindness* 97, 8 (2003), 475–488.
26. Metro Transit: King County Department of Transportation. Stop Information for E Thomas St & 16th Ave E. 2013. <http://goo.gl/aItiY>
27. Noor, M.Z.H., Ismail, I., and Saaid, M.F. Bus detection device for the blind using RFID application. *Proc of CSPA'09*, 247–249.
28. Panciera, K., Priedhorsky, R., Erickson, T., and Terveen, L. Lurking? cyclopaths?: a quantitative lifecycle analysis of user behavior in a geowiki. *Proc of CHI'10*, 1917–1926.
29. Prasain, S. StopFinder: improving the experience of blind public transit riders with crowdsourcing. *ASSETS'11 Poster*, 323–324.
30. Rundle, A.G., Bader, M.D.M., Richards, C.A., Neckerman, K.M., and Teitler, J.O. Using Google Street View to audit neighborhood environments. *Amer. J. of Preventive Medicine* 40, 1 (2011), 94–100.
31. Steinfeld, A., Zimmerman, J., Tomasic, A., Yoo, D., and Aziz, R. Mobile Transit Information from Universal Design and Crowdsourcing. *J of the TRB* 2217, (2011), 95–102.
32. The National Federation of the Blind. Resolutions Adopted by the Annual Convention of the National Federation of the Blind. *The Braille Monitor*, October (1986).
33. Yoo, D., Zimmerman, J., Steinfeld, A., and Tomasic, A. Understanding the space for co-design in riders’ interactions with a transit service. *Proc of CHI'10*, 1797–1806.