tvst

Article

Data Acquisition, Processing, and Reduction for Home-Use Trial of a Wearable Video Camera-Based Mobility Aid

Shrinivas Pundlik[1,2], Vilte Baliutaviciute[1,2], Mojtaba Moharrer[1,2], Alex R. Bowers[1,2], and Gang Luo[1,2]

[1] Schepens Eye Research Institute of Mass Eye & Ear, Boston, MA, USA
[2] Department of Ophthalmology, Harvard Medical School, Boston, MA, USA

Purpose: Evaluating mobility aids in naturalistic conditions across many days is challenging owing to the sheer amount of data and hard-to-control environments. For a wearable video camera-based collision warning device, we present the methodology for acquisition, reduction, review, and coding of video data for quantitative analyses of mobility outcomes in blind and visually impaired participants.

Methods: Scene videos along with collision detection information were obtained from a chest-mounted collision warning device during daily use of the device. The recorded data were analyzed after use. Collision risk events flagged by the device were manually reviewed and coded using a detailed annotation protocol by two independent masked reviewers. Data reduction was achieved by predicting agreements between reviewers based on a machine learning algorithm. Thus, only those events for which disagreements were predicted would be reviewed by the second reviewer. Finally, the ultimate disagreements were resolved via consensus, and mobility-related outcome measures such as percentage of body contacts were obtained.

Results: There were 38 hours of device use from 10 participants that were reviewed by both reviewers, with an agreement level of 0.66 for body contacts. The machine learning algorithm trained on 2714 events correctly predicted 90.5% of disagreements. For another 1943 events, the trained model successfully predicted 82% of disagreements, resulting in 81% data reduction.

Conclusions: The feasibility of mobility aid evaluation based on a large volume of naturalistic data is demonstrated. Machine learning–based disagreement prediction can lead to data reduction.

Translational Relevance: These methods provide a template for determining the real-world benefit of a mobility aid.

Introduction

Vision impairments have been associated with overall decreased mobility and with an increased risk of collisions and falls.[1–5] Mobility-related deficits reported in the literature are predominantly either self-reported via questionnaires and surveys,[6,7] or observed in studies in controlled environments featuring mobility courses,[1] including device/intervention evaluation studies.[8–10] Only certain aspects of mobility can be measured in controlled environments; for example, contacts with obstacles,[8,11] walking speed,[9] object recognition distance,[12] or street-crossing performance,[13] among others. Such studies have a limitation; it is unknown whether mobility deficits associated with visual impairments that are self-reported or observed in constrained, artificial environments accurately represent the mobility challenges during daily activities in a natural environment,

translational vision science & technology

including home, work, outdoors, stores, and other environments.

Some naturalistic walking studies indeed measured real-world mobility in people with visual impairments.[14–16] However, those studies only monitored mobility at a high level via measures such as step counts and/or number of serious falls over a period of time. These studies primarily relied on motion sensors (accelerometers and gyroscopes) and/or GPS sensors to obtain objective mobility data such as step counts[14] and the number of trips made away from home.[17] Motion sensors in some studies also indicated whether there was a fall experienced by the users.[18] Falls can be somewhat easily detected because sensor signals during normal walking can be distinguished from those associated with the fall events. However, fall events are rare and therefore data related to falls are relatively difficult to obtain and require recording for very long periods of time. Usually, motion sensors used for fall detection cannot reliably detect situations where visually impaired users bump into obstacles while walking. Moreover, the nature of the hazard and other relevant factors (environmental conditions during the walk) are not captured by these sensors. Wearable cameras provide an opportunity to obtain rich information about mobility-related challenges, such as collisions with obstacles, along with a more detailed description of the operating environment that can be helpful in providing a more realistic assessment of mobility.

We had previously developed a video camera-based wearable collision warning device as a mobility aid for blind and visually impaired individuals.[11,19] We are conducting a home use trial of the device where the study participants wear the device during their daily activities over multiple weeks, both indoors and outdoors. With this device, we can record video data during device use to provide information about the naturalistic mobility of the users in unconstrained environments. This study is the first we know of that is attempting to investigate collision incidents in naturalistic walking using video cameras. One of the challenges is the sheer volume of video data that are collected and need to be parsed to extract relevant mobility-related information. Currently, there are no known established methods for doing this type of analysis for walking mobility.

Even though there are no established methods for obtaining quantifiable outcomes from naturalistic walking video data in the field of walking mobility, we can borrow some concepts from naturalistic driving research, where driving behavior related outcomes have been obtained from the video data captured by in-car cameras and sensors in the participants' cars.[20]

Our goal is to establish methods for obtaining mobility-related data from naturalistic walking videos captured by a wearable camera, specifically determining the contacts with surrounding objects and categorization of the objects as collision hazards. Such quantitative mobility outcome measures can be recorded by experimenters observing participant's mobility along a predefined indoor or outdoor route in a laboratory study, but that is not possible for home use studies. An intensive manual review is required for annotating the naturalistic videos. Typically, it will take much longer to review and extract information from a video to annotate the details and categorize each mobility event than the actual length of the video. Therefore, our aim is to develop an accurate, objective, and feasible scheme for review and analysis of the naturalistic walking video data.

The objectivity of the outcome measures needs to be maintained by using multiple independent reviewers, given that there is an element of subjectivity in manual video review. Accuracy refers to the ability to obtain specific mobility-related information unambiguously from the video data, such as body contacts with obstacles. Feasibility is an important consideration because a review of all video data may be practically infeasible, and methods for efficient data reduction have to be devised without affecting the overall accuracy of the outcomes.

This article describes the data acquisition scheme, the bases for data review and annotation, along with the formal definitions of each event annotation category or item, and a novel approach for data reduction using machine learning to predict disagreement between the independent reviewers, using the previously known review patterns.

Methods

Naturalistic Walking Data Acquisition

Data acquisition was conducted in the context of a double-masked, randomized controlled clinical trial (NCT03057496) of a wearable video camera based collision warning device that we had previously developed for blind and visually impaired individuals.[21] The study followed the tenets of the Declaration of Helsinki and informed consent was obtained from all the study participants. The protocol was approved by the institutional review board at the Massachusetts Eye and Ear Infirmary and the U.S. Army Medical Research and Materiel Command, Office of Research Protections, Human Research Protection Office.

Data reported in this article are from clinical trial participants with either total blindness or ultralow vision who were all independent travelers and used a long cane or guide dog as their habitual mobility aid. The collision warning device was used in conjunction with their habitual mobility aid. In our overall study sample of 33 participants for the clinical trial, 28 participants reported used a long cane as their primary habitual mobility aid, three participants reported using a guide dog, and two participants indicated that they used both a long cane and a guide dog. The data reported for this manuscript were randomly sampled from 10 of the 33 participants, including nine participants who only used a long cane and one participant who used a guide dog. For the purpose of video review, we did not differentiate between these two mobility aids because the overwhelming majority of events involved a long cane as the mobility aid and the main goal was to determine whether or not a body contact occurred with a hazard after a valid collision warning.

The device camera sensed the environment, computed collision risk, and gave simple directional warnings of collision hazards to the users via vibrotactile wristbands only when collision risk was high (exceeded a predefined, time-to-collision threshold). The goal of the clinical trial was to determine the mobility benefit of the device in the users' daily life activities. Therefore, the study participants used the device over a period of 4 weeks in their everyday mobility.

The device switched intermittently between active mode (providing vibrotactile warnings for detected hazards) and silent mode (hazards detected but no warnings given) in a random manner. The schedule of switching and the duration for which the device remained in each mode varied. The silent mode was the control condition for the clinical trial. Participants, study staff and video reviewers were masked, that is, whether the device was in active or silent mode was unknown to the participants when they used the device, and to the study staff when they reviewed the videos. Although it is crucial for evaluation of the device in the clinical trial, device operating mode is incidental in the context of this article, which focuses on the development of methodology for data acquisition from the videos, data reduction, and development of mobility-related outcome measures.

In its physical form, the device was incorporated within a single strap travel bag, with the video camera situated approximately on the center of the chest, which had a field of view of about 90° horizontally and 60° vertically, covering the head and chest level hazards typically not detected by a long cane (see Pundlik et al.[11] for details regarding the device). Along with
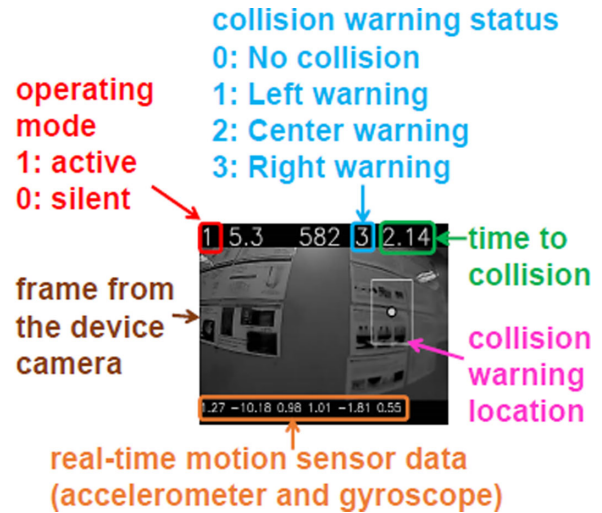

Figure 1. Data recorded by the collision warning device. The chest-mounted video camera captures scene videos, and each video frame is embedded with relevant device data including whether a collision warning was provided, the direction of collision warning (left, center, right), device operating mode, and the real-time motion sensor data. If a collision warning is provided, its location is indicated on the video frame (white box with a dot in the center). This helps in determining the object for which the warning was provided. The text information embedded at the top and bottom of the video frames are extracted by OCR processing, for computerized preprocessing, but they are not visible to study staff in video reviewing.

sensing, the chest-mounted video camera also recorded scene videos during use, thus providing a log of the mobility events encountered by the users.

The device also recorded instantaneous device status information, including whether a collision warning was provided and, if so, the location of the collision warning in the current video frame (denoted by a box with a dot in the center). These device data were embedded into the scene video frame and were therefore a part of the recorded videos (Fig. 1). Embedding device data as text within video frames allowed for easier synchronization between the device action and the scene video. For example, when a collision warning was provided to the user, it was logged within the video and the reviewer could view and analyze the marked video segments to see where and why the collision warning was provided. Throughout use, the device recorded these videos (grayscale, 320 × 288 resolution) and stored them on a memory card. After use, the video data from the memory card was transferred to desktop computers for further processing.

Data Processing

After the walking videos were obtained, we extracted the embedded text data, detected mobility events of interest, and masked the videos to prepare
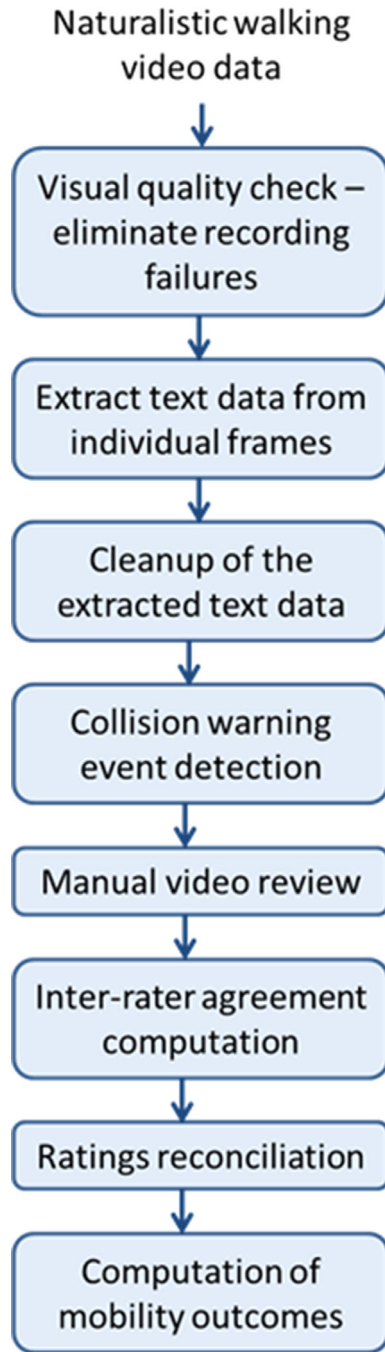
Naturalistic walking
video data

↓

Visual quality check –
eliminate recording
failures

↓

Extract text data from
individual frames

↓

Cleanup of the
extracted text data

↓

Collision warning
event detection

↓

Manual video review

↓

Inter-rater agreement
computation

↓

Ratings reconciliation

↓

Computation of
mobility outcomes

Figure 2. Flowchart showing the steps in video data processing to obtain quantifiable mobility outcomes.

for video review. The steps involved in this operation are shown as a flowchart in Figure 2. Video icons were visually inspected to check for valid data, so that occasional recording failures (black screen) could be eliminated from further processing. Each video was then processed frame by frame. The top and bottom strips containing text data were cropped and the video part saved for further viewing, to ensure that the reviewers were masked to device status while reviewing the videos. An optical character recognition (OCR) software routine[22] processed the top and bottom strips of the frames to extract the device status and motion sensor information, respectively. These extracted data were stored in text files (a text file for a video contains frame-by-frame information).

The OCR software made occasional transcription mistakes, particularly because the videos were low resolution and occasionally suffered from compression artifacts. Thus, there was a possibility that data for certain frames could be garbled, which needed to be either corrected or eliminated. A follow-up software routine was run on the extracted text data to detect and wherever possible, correct the OCR mistakes. Because the format of the text data, their location within the video frame, and the expected ranges of the values within each field were known, error correction could be done to recover most of the text data. The most common mistakes were missing spaces, and with the known text format, those could be corrected. Missing or seriously garbled text data were eliminated. The entire process of extraction of text from video frame along with OCR error correction was automated.

After cleaning up the text data obtained via the OCR software, collision warning event detection was performed. The device provided collision warnings on a per-frame basis, which means either the given frame had a collision warning or did not. In actual use, a collision threat could unfold over a span of multiple video frames. For example, with the participant approaching an obstacle, the device could provide warnings over a short duration of time on the order of a few seconds. To make review consistent and feasible, all collision warnings within a span of 2 seconds were grouped as a single event. This time window of 2 seconds for grouping the collision warnings was chosen empirically. Once all the collision risk events were computed within a video, further processing and review was done with reference to these events rather than the video frames. The event identification process within the recorded videos was automated.

Video Review

Manual review of the detected collision warning events was required to determine why the device gave warnings and what really happened when the warnings were given. The detected events and the corresponding scene video (devoid of embedded text information) were fed to custom video review software for manual inspection. Reviewers could move from one event to another and play a short video clip around the detected event to annotate the relevant event details. Event details such as whether there was a
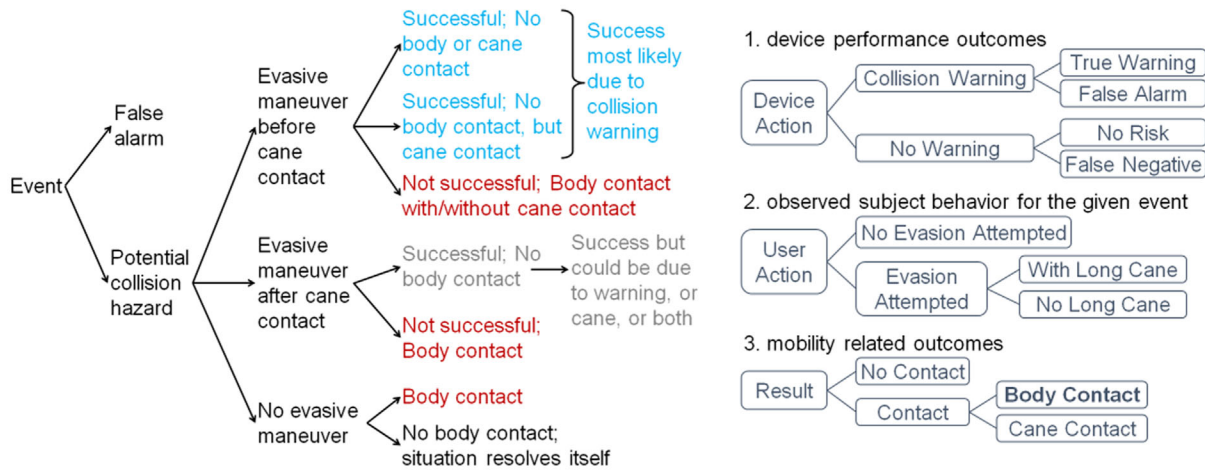
Figure 3. Reviewing and coding a collision warning event. (*Left*) An event can unfold in a complex manner, and depending on how it unfolds and the action taken by the user could result in contact with the obstacle. Following a complex tree for detailed annotation of an event may not be feasible or possible directly via video review. Success and failure can be either defined from a user's perspective or from device's perspective. From user's perspective, not having a body contact can be considered as a success, irrespective of the reason. From device's perspective, a cane contact may be considered a failure even if there is no body contact, depending on when the cane contact happens. (*Right*) Conceptually breaking down an event into three categories: device performance, user action, and the final result, can help to simplify the coding of an event while maintaining thoroughness of the review process.

collision hazard, whether there was any contact with the hazard, the nature of the hazard, and the nature of the scene/location where the collision hazard was observed (whether the collision hazard was in participants' familiar environment [home/office] or not), were annotated.

The main goal of event annotation was to obtain quantifiable mobility measures from video observation. The main mobility-related outcome of interest was the number of body contacts with detected hazards. Other relevant mobility-related data included the number of cane contacts, the number of true hazards encountered, the nature of the collision hazard, and the walking environment, among others. Just considering the mobility-related outcomes, each event can unfold in various different ways leading to a complex flow diagram (Fig. 3, left) because there are many interdependent steps between the first stage of the device issuing a warning to the final outcome (contact or no contact). Annotating these details is difficult just based on the video captured from the chest-mounted camera. Therefore, to simplify and streamline the review process, the event-related details that needed to be annotated were classified into the following broad categories: device action (whether it was true hazard or false alarm), user action (what did the user do), event outcome (whether there was a body contact, cane contact, or none), and the environment (Fig. 3, right). This process resulted in a hierarchical review flowchart, where certain quantities such as body contact depended on whether there was a contact of

any kind, including contact with long cane, which in turn depended on whether there was a true hazard. Given that most of the events involved a long cane as the mobility aid, any contacts with mobility aids are generally referred to as cane contacts in the text of this article.

Even after further simplification in reviewing categories, it may not always be possible to accurately annotate the details in an event just based on the video. For example, in certain cases it might not be possible to tell whether the participant hit an object with their cane because the end of the cane might not be within the field of view of the video. Similarly, in many other situations the action of the participant as well as the outcome of the event may not be obvious and therefore subjective judgment could lead to arbitrary outcomes. To address this issue, we first drafted formal definitions of all the event annotation categories based on observable evidence that would help in the subjective judgment. The formal definitions were based on preliminary scoring of 338 events by authors SP, VB, and MM. The definitions were then refined through an iterative process involving all authors in which unambiguous and ambiguous events of various types were reviewed in a group setting and possible interpretations discussed until consensus was reached (Table 1). After developing the definitions, we implemented a reviewing scheme involving two masked reviewers (VB and MM) independently reviewing the data to further improve the objectivity of the review process.

Table 1.   Definitions of the Annotation Categories Used to Rate Events

Annotation Category	Options	Meaning
Valid event[a]	Yes/no	Yes: Camera view was unobstructed, device operation as expected.
		No: Device operation was disrupted in some way. E.g. user hand obstructed camera, light glare created a visual artifact, or the device is not being worn.
True hazard[a]	Yes/no	Yes: The warning was valid and associated with a true hazard; a collision would have occurred if the trajectory of motion was maintained.
		No: the warning was a false alarm.
Evasion attempt[a]	Yes (cane not involved, cane involved, not sure)/no	Yes (cane not involved): There was an evasion attempt (e.g. step to the side) with no clear use of long cane.
		Yes (cane involved): There was an evasion attempt after cane contact with the obstacle.
		Yes (not sure): There was an evasion attempt, but it is unclear whether the long cane made contact. Use sparingly.
		No: There was no visible evasion attempt.
Contact[a] (all contact/body contact)	Cane contact/body contact/not sure/no	Cane contact: Participant made contact with the obstacle with their habitual mobility aid (long cane or the guide dog). In the absence of direct visual evidence, contact could be inferred by a sudden pause, sharp change of walking direction, or jolting/shaking of the camera, together with the relative distance to the object in the scene.
		NOTE: This option was also considered when a participant used their hand to find an obstacle that they were aware of.
		Body contact: Participant collided with the obstacle directly. Notable by more severe camera jolt and close camera view. If both cane and body contact occur, mark as body.
		Not sure: a contact occurs, but it is ambiguous whether with cane or body. Use sparingly.
		No: there was no contact of any kind.
Home/office vs. other environment	Yes/no	Yes: The scene is inside participant's home/work environment.
		No: The scene is outside participant's home/work related environment, such as streets, shopping mall, or transit stations, etc.
Nature of the hazard	Pedestrian, furniture, poles, walls, overhanging, trees, other	Pedestrian: the hazard was a person
		Furniture: desks, chairs, shelves, racks, etc.
		Poles: Poles, posts, pillars, bollards, columns, other similar standing structures.
		Walls: Walls, doors, building structures, etc.
		Overhanging: Tree branches, flags, banners and similar hanging/head-height objects.
		Trees: Tree trunks, bushes, hedges, etc.
		Other: Anything that doesn't fit the above categories (lights, vehicles, etc.)

Table 1. Continued

Annotation Category	Options	Meaning
Moving camera	Yes/no	Yes: The user was in motion (walking, swaying, on an escalator). No: the user is still (sitting/standing).
Moving object/hazard	Yes/no	Yes: the hazard is moving (e.g., a walking pedestrian) No: The hazard is still (e.g., stationary furniture)
Left turn	Yes/no	This is only selectable if there is an evasion attempt, and notes the direction of the evasion.
Right turn	Yes/no	This is only selectable if there is an evasion attempt, and notes the direction of the evasion.

[a]The categories of valid event, true hazard, evasion attempt, and contact are critical for assessing mobility outcomes and device performance. Only these categories were relevant for disagreement reconciliation. The other categories provide additional detail, such as what the hazard was, or where the user was at the time of the collision hazard event.

The home use trial data for a given participant consisted of multiple short videos (maximum duration of 15 minutes; longer videos were broken down into 15-minute segments by the video recorder). Each video could contain a different number of events (some had no events detected). For reviewing, the video order for a given participant was randomized, but the events occurring in the same video were not randomized. For the data presented in this article, events were reviewed by both the reviewers independently and then the annotations were compared to determine disagreements. Disagreements between the reviewers were reconciled with consensus for the following review categories: valid event, true hazard, all contacts, and body contacts. These four items were important in our study for determining the mobility-related outcomes for naturalistic walking. They were coded hierarchically: first whether the event was valid, then whether it was a true hazard if it was valid, then whether there was any kind of contact if it was a true hazard, and finally whether there was a body contact if there was a contact. The probability of agreement and Cohen's kappa values were computed to provide inter-rater reliability metrics between the two reviewers for these four categories.

Data Reduction

The feasibility of data review is a big concern because a large amount of video data requires a lot of manual effort. In particular, when multiple reviewers review the same data, the total effort level becomes even higher. However, multiple independent reviewers are needed for maintaining the objectivity of the assessments. Therefore, techniques for data reduction had to be devised for making reviewing and coding feasible to

perform. Data reduction here refers to the duration of video data that needs to be manually reviewed relative to the overall duration, and therefore larger data reduction is preferable for the feasibility of manual review, as long as we do not eliminate relevant events as part of the data reduction process. One obvious way of data reduction, which is inherent in the method we implemented, was reviewing only the segments of videos where the device provided collision warnings. Thus, the event-driven review cut down on the overall time and we could avoid reviewing the entire videos at full length. However, this data reduction was still not sufficient owing to the large number of events detected by the device.

To further decrease the reviewing effort, we focused on a novel strategy to predict disagreements between the two reviewers based on how they previously rated the same events. If we could predict events where the two reviewers were likely to disagree, then each reviewer would only have to review a subset of the entire data, thus saving on time and effort. So, in this novel scheme, the two reviewers look at different events in the initial round. Then, based on the events that have been reviewed by both the reviewers previously (reviewing history), we can predict where they might disagree. Then, they swap the events with each other and only review those that they were predicted to disagree with. In this manner, the amount of data that they were expected to review can be substantially decreased while maintaining the accuracy and objectivity of the outcomes.

To predict events where the reviewers might disagree, we used RUSBoost classifier[23,24] implemented in the MATLAB Classification Learner App. Training data consisted of each individual reviewer's coding of multiple events across 11 different items: valid event, true hazard, all contacts, body contact, left

turn, right turn, evasion attempt (all causes), evasion attempts—cane not involved, moving camera, moving object/hazard, and the scene settings (home/office vs. others). Both the reviewers were highly trained before on separate video data (not used here). After reviewing the same data independently, disagreements for different review items were obtained. These known disagreements in the review of body contacts (our primary mobility outcome) were the labeled output corresponding to the rest of the review items, and together they constituted the training data. The classifier was trained on the data belonging to each reviewer separately, to recognize the patterns of ratings in these 11 review items that were more likely to lead to a disagreement about body contact for an event. The disagreement prediction algorithm was tuned to decrease the false-negative rate (the proportion of events where the algorithm did not predict a disagreement on body contact when it should have). Automated feature selection was used to retain predictors that contributed significantly to the overall model at 95% confidence. A five-fold cross-validation scheme was used for evaluation.

Implementation of the Review Scheme

First, data processing software was developed in Matlab to automate gathering of collision warning event data from the recorded videos (steps shown in Fig. 2). Then, preliminary event scoring criteria were conceived and a custom review software was developed that allowed playback and annotation/review (via check boxes and drop down menus) of individual events within a video. The software could jump back and forth between events within a video. The initial training of the reviewers and refining of the review criteria were performed iteratively, with the reviewers viewing the same videos during pilot stages of the study and then reconciling differences in review in joint meetings with all study investigators. At the same time, the reviewers' inputs regarding which scoring items were feasible and important were incorporated into the review software. Video data collected from visually impaired and blind participants during pilot testing of the device in habitual mobility were used during the development of the review criteria. Once the review criteria were finalized, the two reviewers then independently reviewed a large number of events from data collected in the early part of the clinical trial, and these data were used to train the machine learning algorithm to predict events where the two reviewers might disagree on whether there was a body contact.

Figure 4. Agreement/disagreement between the 2 masked reviewers when performing manual review of the video data. A total of 2712 events were reviewed independently by each reviewer (rater A and rater B). The four review items shown here were rated hierarchically in following order: valid event, true hazard, all contacts, and body contacts. If both the reviewers rated no for any given item, the event was dropped from consideration for subsequent review items. Therefore, the total number of events was lower for items lower in the hierarchy.

Results

A total of approximately 38 hours of device use video data across 10 blind or visually impaired participants were selected for analysis for this study. Text extraction with the OCR engine was largely successful, with only 0.35% of all the video frames returning no text data (success rate of 99.65%). Automated processing of the extracted text data from the video frames revealed a total of 2712 collision warning events registered by the device. Detailed annotation of each event separately performed by the two independent reviewers was compiled. This event review for 2712 events by both the reviewers along with their disagreements regarding body contacts served as the training data for the machine learning algorithm for disagreement prediction.

Figure 4 shows the 2 × 2 agreement tables for the four main review items between the two reviewers over all 2712 events after the initial round of review (before reconciliation). Because these items were rated hierarchically, the events where both the reviewers answered in negative were not considered in the subsequent items at the lower hierarchy levels. Therefore, the total number of events in the tables for true hazard, all contacts, and body contacts progressively decreased. Agreement probabilities and Cohen's kappa for the four items are shown in Table 2. The reviewers concurred most (96% of events) for the valid event category and concurred least (66% of events) for body contacts. The Cohen's kappa values ranged from 0.67 (valid event) to 0.05 (for body contacts).

Figure 5 shows the confusion matrices for disagreement prediction related to body contacts for the two reviewers after five-fold cross-validation with 2712 labelled event samples. For the 2712 events rated by rater A, the algorithm correctly predicted 176 out of the total 200 already identified disagreements (Fig. 4, far right), a success rate of 88%. For the same events rated

Table 2. Inter-Rater Reliability Between the 2 Independent Reviewers for Ratings of Valid Event, True Hazard, all Contact, and Body Contact Across 2712 Events

Measure[*]	Agree	Disagree	Agreement Probability	Cohen's Kappa
Valid Event	2592	120	0.96	0.67
True Hazard	2035	539	0.79	0.57
All Contacts	902	428	0.68	0.24
Body Contacts	391	200	0.66	0.05

[*]The order of listing of the items in the table represent the hierarchy followed when scoring these items for a given event. Therefore, the total events reduce as we move from valid event to body contact ratings.
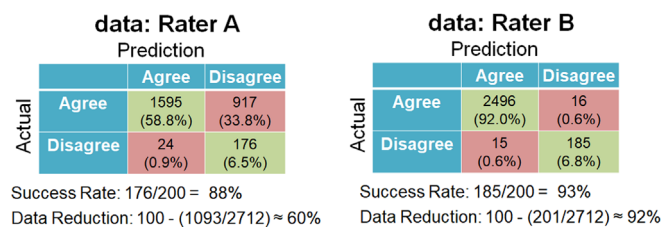

Figure 5. Results for predicting disagreements in rating of body contacts during event review by the two raters. The machine learning algorithm was trained on each reviewer's ratings for the same 2712 events with 200 known disagreements. The % values in the table are relative to the total events reviewed (2712). Results were computed using five-fold cross-validation for these set of events. For data reviewed by rater A, the algorithm predicted 176 disagreements, with rater B correctly, while missing 24 (success rate of 88%). For data reviewed by rater B, the algorithm predicted 185 disagreements with rater A, while missing 15 (success rate of 93%).

by Rater B, the algorithm predicted 185 out of 200 disagreements (success rate of 93%). The total number of disagreements predicted with data reviewed by rater A were 1093, amounting to a data reduction of about 60%. For the data reviewed by rater B, the total number of disagreements predicted by the algorithm was 201, and the data reduction was at about 92%.

In a further test of the algorithm, a new dataset, which was not previously used in training the machine learning algorithm, was fed to it to predict the disagreements in body contacts. For the 1943 events reviewed first by rater A, the algorithm predicted body contact disagreement for 511 events. After review by rater B, actual disagreements were found to be 25 (with 100% overlap between actual disagreements and algorithm prediction) and an overall data reduction of approximately 74%. For a separate set of 1875 events reviewed by rater B, the algorithm only predicted disagreements in body contact for 35 events. Actual disagreements were 34, and the algorithm predicted 25 of the 34 disagreements (success rate of 74%). The data reduction in this case was approximately 98%. On average, the algorithm could predict disagreements between

two reviewers with 82% success rate, with an average data reduction rate of 81%.

Discussion

The approach described in this article provides a blueprint to tackle challenging big data analysis problems related to collisions in daily mobility of visually impaired and blind participants. The main contributions of our approach are (i) applying robust methods for quantification of mobility related outcomes from video data recordings in the daily mobility of people with severe visual impairments, and (ii) proposing a novel algorithm for data reduction to make the analysis effort feasible.

Our approach focuses on the previously unaddressed issue of analyzing large amounts of video data to obtain mobility-related outcome measures relevant to the use of devices to assist in obstacle detection and collision avoidance when walking. Previous studies about naturalistic walking mobility in visually impaired individuals mainly analyzed motion sensor data (number of steps and/or falls) and primarily focused on a particular group of patients or disease category (such as glaucoma[14,17,18,25] or AMD[15,26]), where the collision risk was presumably lower compared with people with more severe visual impairments or blindness who were the focus of our study. Although the proposed methods were designed and tested for data involving blind or severely visually impaired individuals, the same methods could be used when investigating real-world mobility in other patient populations.

The inter-rater reliability varied between different review items, with classification of valid events being the highest, followed by true hazard, all contacts, and body contacts. In other words, it was easier to tell whether an event was valid or not than to tell whether there was a body contact. Given the wide variability between the scenarios where the events took place, it

is conceivable that no matter how closely aligned the two raters are, there will be disagreements when classifying for body contacts. Therefore, multiple independent reviews followed by consensus based reconciliation can ensure that the most important outcome measure is obtained with relatively high reliability despite disagreements.

The data reduction technique was designed with the same goal of obtaining important mobility-related outcomes with high reliability. The disagreement prediction algorithm was tuned to ensure most potential disagreements were not missed, possibly at the cost of an increased false alarm rate (predicting disagreement for an event when there was no disagreement). Failing to quantify a body collision has negative consequences for data analyses. False alarms increase the amount of data that need to be reviewed but, as our study showed, the algorithm predictions covered about 82% of the disagreements in the body contact rating and greatly decreased the number of events that needed to be reviewed by both reviewers (by 81%).

The two raters exhibited differing categorization patterns when reviewing the data. These two individual reviewing patterns were used to train the disagreement prediction algorithm. Based on the review of events by rater B, it was relatively easy to determine which events rater A would disagree with in terms of body contact. However, the opposite was not necessarily true for the data reported here.

Once trained on a common set of data reviewed fully by two individuals, the algorithm should work as long as the same two individuals continue to do all the reviewing. However, if a new pair of reviewers is to be inserted, then they both will have to review a common set of events in sufficient numbers for the machine learning algorithm to learn their reviewing patterns. In our case, when training the algorithm, we worked with a sample of 2712 common events that were reviewed by both reviewers. Considering each event takes on average 1 minute to review (but new reviewers might take longer than trained reviewers), the lead time to retrain the disagreement prediction algorithm could be about 45 hours of reviewing per reviewer (90 hours for a new pair of reviewers). After the algorithm has been trained, depending on the algorithm performance, we can expect significant savings in the reviewing efforts compared to full double reviewing of all events by both reviewers. To put these savings in context, consider the data set from the clinical trial which currently consists of more than 29,000 events (at least 483 hours of reviewing for each reviewer). Initial, full double reviewing needs to be done only for about 10% of the total events for training the algorithm. For the remaining 90% of the data, the reviewing effort reduction will

be substantial, on average 80%, resulting in approximately 12 fewer hours per thousand events reviewed. The reviewing effort reduction will likely vary between pairs of reviewers and could be more or less than found for the two reviewers in this study. Nevertheless, we suggest that a data reduction of 80% is a realistic expectation given that our two reviewers exhibited clearly different categorization patterns when reviewing.

Possible alternatives to the presented approach of video review might include crowdsourcing and artificial intelligence approaches. Crowdsourcing can be an efficient way to save researchers' effort, particularly for relatively simple tasks such as image labeling, but may not be feasible for complex tasks such as detailed mobility video annotation that require nontrivial user training. Given the complexities of obstacle avoidance when walking in the real world, the reviewers for our particular application need to be aware of the functionality and limitations of the device. Also, there is little control over who reviews what in crowdsourcing, and therefore reconciliation of disagreements is not as straightforward as in our approach (joint review of items with disagreements). Another alternative approach, based on artificial intelligence algorithms to automatically review and annotate events, holds promise for future work.

In conclusion, our novel approach resulted in a data reduction of about 80%, which means that the actual amount of video to be reviewed will only be 19% of the original data. For the first time, our approach makes it possible to objectively study and quantify collision incidents in daily mobility of visually impaired and blind individuals, and makes it feasible to conduct clinical trials to objectively evaluate the effectiveness of video camera-based mobility assistance devices in habitual mobility. Furthermore, the approach described in this article may be helpful in providing a better understanding of the processes involved in and difficulties encountered during obstacle detection and avoidance when walking.

Acknowledgments

Disclosure: **S. Pundlik**, (P); **V. Baliutaviciute**, None; **M. Moharrer**, None; **A.R. Bowers**, None; **G. Luo**, (P)

References

1. Turano KA, Borman AT, Bandeen Roche K, Muñoz B, Rubin GS, West SK. Association of visual field loss and mobility performance in older adults: Salisbury Eye Evaluation Study. *Optom Vis Sci*. 2004;81:298–307.

2. Freeman EE, Muñoz B, Rubin G, West SK. Visual field loss increases the risk of falls in older adults: the Salisbury Eye Evaluation. *Invest Ophthalmol Vis Sci*. 2007;48:4445–4450.

3. Dhital A, Pey T, Stanford M. Visual loss and falls: a review. *Eye*. 2010;24:1437–1446.

4. Lovie-Kitchin JE, Soong GP, Hassan SE, Woods RL. Visual field size criteria for mobility rehabilitation referral. *Optom Vis Sci*. 2010;87:948–957.

5. Taylor DJ, Hobby AE, Binns AM, Crabb DP. How does age-related macular degeneration affect real-world visual ability and quality of life? A systematic review. *BMJ Open*. 2016;6:e011504.

6. Manduchi R, Kurniawan S. Mobility Related Accidents Experience by People with Visual Impairment. *Insight: Research and Practice in Visual Impairment and Blindness*. 2011;4:1–11.

7. Soubrane G, Cruess A, Lotery A, et al. Burden and health care resource utilization in neovascular age-related macular degeneration: findings of a multicountry study. *Arch Ophthalmol*. 2007;125:1249–1254.

8. Pundlik S, Tomasi M, Luo G. Evaluation of a portable collision warning device for patients with peripheral vision loss in an obstacle course. *Invest Ophthalmol Vis Sci*. 2015;56:2571–2579.

9. Bowers AR, Luo G, Rensing NM, Peli E. Evaluation of a prototype minified augmented-view device for patients with impaired night vision. *Ophthalmic Physiol Opt*. 2004;24:296–312.

10. Geruschat D, Bittner AK, Dagnelie G. Orientation and mobility assessment in retinal prosthetic clinical trials. *Optom Vis Sci*. 2012;89:1308–1315.

11. Pundlik S, Tomasi M, Moharrer M, Bowers AR, Luo G. Preliminary evaluation of a wearable camera-based collision warning device for blind individuals. *Optom Vis Sci*. 2018;95:747–756.

12. Zebehazy KT, Zimmerman GJ, Bowers AR, Luo G, Peli E. Establishing mobility measures to assess the effectiveness of night vision devices: results of a pilot study. *J Vis Impair Blind*. 2005;99:663–670.

13. Bowman E, Liu L. Individuals with severely impaired vision can learn useful orientation and mobility skills in virtual streets and can use them to improve real street safety. *PLoS One*. 2017;12:e0176534.

14. Ramulu PY, Maul E, Hochberg C, Chan ES, Ferrucci L, Friedman DS. Real-world assessment of physical activity in glaucoma using an accelerometer. *Ophthalmology*. 2012;119:1159–1166.

15. Curriero FC, Pinchoff J, van Landingham SW, Ferrucci L, Friedman DS, Ramulu PY. Alteration of travel patterns with vision loss from glaucoma and macular degeneration. *JAMA Ophthalmol*. 2013;131:140–1426.

16. van Landingham SW, Willis JR, Vitale S, Ramulu PY. Visual field loss and accelerometer-measured physical activity in the United States. *Ophthalmology*. 2012;119:2486–2492.

17. Ramulu PY, Hochberg C, Maul EA, Chan ES, Ferrucci L, Friedman DS. Glaucomatous Visual Field Loss Associated with Less Travel from Home. *Optom Vis Sci*. 2014;91:187–193.

18. Ramulu PY, Mihailovic A, West SK, Gitlin LN, Friedman DS. Predictors of falls per step and falls per year at and away from home in glaucoma. *Am J Ophthalmol*. 2019;200:169–178.

19. Pundlik S, Tomasi M, Luo G. Collision detection for visually impaired from a body-mounted camera. *Proc IEEE Conf Comp Vis Pattern Rec Workshops*, 2013;41–47.

20. Dingus TA, Guo F, Lee S, et al. Naturalistic driving evaluation of crash risk. *Proc Natl Acad Sci USA*. 2016;113:2636–2641.

21. ClinicalTrials.gov. Collision warning device for blind and visually impaired. Available at: https://clinicaltrials.gov/ct2/show/NCT03057496?term=NCT03057496&draw=2&rank=1. Accessed 2017.

22. Tesseract OCR. Available at: https://github.com/tesseract-ocr/tesseract. Accessed June 2, 2020.

23. Seiffert C, Khoshgoftaar TM, Hulse JV, Napolitano A. RUSBoost: improving classification performance when training data is skewed. *International Conference on Pattern Recognition*. Tampa, FL: IEEE Computer Society; 2008.

24. Seiffert C, Khoshgoftaar TM, Hulse JV, Napolitano A. RUSBoost: a hybrid approach to alleviating class imbalance. *IEEE Trans Syst Man Cybern A Syst Hum*. 2010;40:185–197.

25. Lee MJ, Wang J, Friedman DS, Boland MV, De Moraes CG, Ramulu PY. Greater physical activity is associated with slower visual field loss in glaucoma. *Ophthalmology*. 2019;126:958–964.

26. Sengupta S, Nguyen AM, van Landingham SW, et al. Evaluation of real-world mobility in age-related macular degeneration. *BMC Ophthalmol*. 2015;15:9.