HAL
open science

Personal Shopping Assistance and Navigator System for Visually Impaired People

Paul Chippendale, Valeria Tomaselli, Viviana d'Alto, Giulio Urlini, Carla Maria Modena, Stefano Messelodi, Sebastiano Mauro Strano, Günter Alce, Klas Hermodsson, Mathieu Razafimahazo, et al.

To cite this version:

Paul Chippendale, Valeria Tomaselli, Viviana d'Alto, Giulio Urlini, Carla Maria Modena, et al.. Personal Shopping Assistance and Navigator System for Visually Impaired People. ACVR2014: Second Workshop on Assistive Computer Vision and Robotics, Sep 2014, Zurich, Switzerland. hal-01102707

HAL Id: hal-01102707

https://inria.hal.science/hal-01102707v1

Submitted on 13 Jan 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Personal Shopping Assistance and Navigator System for Visually Impaired People

Paul Chippendale[1], Valeria Tomaselli[2], Viviana D'Alto[2], Giulio Urlini[2], Carla Maria Modena[1], Stefano Messelodi[1], Sebastiano Mauro Strano[2], Günter Alce[3], Klas Hermodsson[3], Mathieu Razafimahazo[4], Thibaud Michel[4], Giovanni Maria Farinella[5]

[1]Fondazione Bruno Kessler, Trento, Italy;
[2]STMicroelectronics, Italy;
[3]Sony Mobile Communications, Lund, Sweden;
[4]Inria Grenoble - Rhône-Alpes/LIG, France;
[5] Image Processing Laboratory (IPLAB), University of Catania, Italy

Abstract. In this paper, a personal assistant and navigator system for visually impaired people will be described. The showcase presented intends to demonstrate how partially sighted people could be aided by the technology in performing an ordinary activity, like going to a mall and moving inside it to find a specific product. We propose an Android application that integrates Pedestrian Dead Reckoning and Computer Vision algorithms, using an off-the-shelf Smartphone connected to a Smartwatch. The detection, recognition and pose estimation of specific objects or features in the scene derive an estimate of user location with sub-meter accuracy when combined with a hardware-sensor pedometer. The proposed prototype interfaces with a user by means of Augmented Reality, exploring a variety of sensorial modalities other than just visual overlay, namely audio and haptic modalities, to create a seamless immersive user experience. The interface and interaction of the preliminary platform have been studied through specific evaluation methods. The feedback gathered will be taken into consideration to further improve the proposed system.

Keywords: Assistive Technology, Indoor Navigation, Visually Impaired, Augmented Reality, Mobile Devices, Wearable Cameras, Quality of Experience

1 Introduction

Data from the World Health Organization [1] reports that 285 million people are visually impaired worldwide: 39 million are blind and 246 million have low vision. Although about 80% of all visual impairments can be avoided or cured, as a result of increasing elderly population more people will be at risk of age-related low vision. In fact, about 65% of all people who are partially sighted are aged 50 and above. In the light of this data, systems that help with navigation

in unfamiliar environments are crucial for increasing the autonomous life of visually impaired people.

We present a hardware/software platform, based on Android OS, integrating outdoor/indoor navigation, visual context awareness and Augmented Reality (AR) to demonstrate how visually impaired people could be aided by new technologies to perform ordinary activity, such as going to a mall to buy a specific product. In particular, indoor navigation poses significant challenges. The major problem is that the signals used by outdoor locating technologies (e.g. GPS) are often inadequate in this environment. Some solutions available exploit the Earth's magnetic field [9], using magnetometers available in Smartphone and relying on maps of magnetic fields; some other solutions rely on the identification of nearby WiFi access points [28], [10], but do not provide sufficient accuracy to discriminate between individual rooms in a building. An alternative approach is Pedestrian Dead Reckoning (PDR) [24], [21]. It composes an inertial navigation system based on step information to estimate the position of the pedestrian. The proposed system integrates a PDR-based system with computer vision algorithms to help people to be independent both indoor and outdoor. The implemented Pedestrian Dead Reckoning uses advanced map-matching algorithms for refining the user's trajectory and also her/his orientation, thus resulting in a significant enhancement of the final estimated position. Computer vision provides nowadays several methods for the recognition of specific objects; users with visual disabilities could benefit from this technology but, unfortunately, the success often relies upon the user being able to point the camera toward the object of interest, that must cover the major part of the image. If the object is embedded in a complex scene and its apparent size is only few pixels, the available recognition apps, usually, fail. The proposed visual target detectors can detect and localize an object of interest inside a complex scene, and also estimate its distance from the user, which is particularly helpful for the guide of a visually impaired person.

The proposed system aims also to provide an AR experience to the user; reality augmentation is not only about visual overlays, but it also encompasses other sensorial modalities. By restricting the modes of AR feedback to the user, i.e. to only non-visual means, the demonstrator leverages on audio and haptic modalities. Differently from the current AR experience which tends to be rather cumbersome and obtrusive to the user, we aim to create a seamless immersive user experience, hiding technology complexity from the user.

Another challenging goal of this project is to deliver pertinent information in a 'user' rather than a 'device' centric way. User experience understanding is essential to make assistive technology really useful, non-obtrusive and pervasive. Assistive technologies for visually impaired people cannot rely solely on design guidelines for traditional user interfaces. Quality of Experience will thus be guaranteed by iteratively performing user studies, to get insights into social acceptance and to receive feedback for further developments.

The remainder of this paper is organized as follows: Section 2 will describe the use case addressed by the proposed system, with reference to a generic story-

board and the actual implementation; in Section 3 all of the exploited hardware and software technologies will be presented; then, in Section 4, the user experience evaluation methodologies will be described, analyzing results and points to be improved; finally conclusions will be drawn in Section 5.

2 Use Case

The use case takes the form of a navigator for visually impaired people, capable of guiding a user through different elements of a mall, until she/he finds a shop to buy a specific product. The implementation of this use case requires a hardware/software platform capable of integrating the hardware-sensor pedometer with algorithms that elaborate visual inputs. Beyond the provision of a location and orientation estimate inside the mall, the system also detects possible hazards in the path of the user and is capable of routing a partially sighted user towards the desired goal. To build such a complex platform, the use case was firstly described by means of a storyboard and then it was broken down into elementary parts, each one representing a section of the path, each characterized by a specific set of algorithms (and problems) for location and navigation, in order to solve them in a tractable manner. The following subsections will describe the storyboard and the real implementation, which has been installed in a large office building to test the whole hardware/software system, which will be called VeDi.

2.1 Storyboard

The addressed use case is focused on Marc who suffers from retinitis pigments, meaning that his visual acuity has been reduced to 20/400 in his best eye. Marc finds it difficult to find exactly what he wants in the shops nowadays. It's his nephew's 12th birthday tomorrow, and he would like to buy him his favourite toy, Lego. He searches on the Internet and finds a good price at the 'Super Shopping Mall' in the Toys 'R Us shop. From 'Google directions' he knows that Bus 73 will take him from his home to the Shopping Mall. He downloads all the maps necessary for the navigation outside and inside the shopping mall onto his VeDi device. After taking the correct bus to the shopping mall Marc arrives at the bus stop nearest to it. Once he steps off the bus, as his VeDi device knew his destination and had pre-downloaded the building plan from OpenStreetMap [2], VeDi guides him to the entrance. The VeDi device drives Marc from the entrance of the mall to the elevator that will bring Marc to the floor of the mall where the toy shop is located. Marc senses there is a potential danger sign in front of him, he stops and asks the VeDi device for confirmation. The VeDi device visually detects the wet floor warning sign and provides an audio alert "walk slowly". VeDi directs Marc to the closest lift, as this is the safest path to the specific shop that VeDi believes Marc needs. VeDi guides him to the buttons using visual cues and then guides Marc's hand to make sure he presses the correct button. As Marc has had normal vision for most of his life, he has a good mental picture

of the environment, but the re-assurances provided by VeDi makes his trip to the mall less stressful. The target floor in this scenario is the second floor. The lift doesn't have aural feedback but VeDi sees that Marc pressed the correct floor button, so Marc gets out of the lift on the correct floor. A lot of people are present in the mall's main thoroughfare and, as a result, VeDi replans a route for Marc that should be quieter. Marc is safely guided to the correct shop and from its entrance to the Lego shelves where he can find his gift. Once in front of the toy shelf, VeDi sees that there are several boxes close together and the platform scans the shelf and detects candidate positions where the desired toy might be. Marc orients his VeDi in front of the shelf until the desired box is found.

2.2 Use Case Implementation

For the sake of simplicity and repeatability, the demo was executed in a large office building, guiding a user from the bus stop, through the main entrance to an office located on another floor via a lift. The use case demo has been divided into several phases, which are listed below:

1. In the first phase the user is guided through the outdoor section of the route from a bus stop to the sliding doors representing the entrance of the mall; a PDR system uses the Smartphone's accelerometers to evaluate the foot-steps of the user, guiding her/him through the path;
2. The second phase guides the user through the main hall towards the lift using only a Visual Navigator, searching for specific visual beacons (like signs or environmental features) in order to identify the optimal direction and frequently making a distance estimate to them;
3. The third phase of the guidance ensures that the visually impaired person correctly uses the lift and its external and internal buttons to the first floor. It visual searches for the button panel and then tracks the user's finger to the correct button to be pressed;
4. The fourth phase restarts the navigation on the first floor and dynamically selectes the best path for the user, through an understanding of whether a corridor is crowded by people or not; a re-route then is offered in order to avoid crowded areas/corridors in favour of freer passages;
5. The fifth phase starts as soon as a free corridor is detected, and a Structure from Motion algorithm is activated for predicting the movement of the user with respect to a starting point. Any drift introduced by the algorithm is compensated by the PDR;
6. The sixth phase starts when the user is in the vicinity of the shop. The toy shop sign over the door is visually recognized as well as the logo LEGO that identifies the target shelves for the demo;
7. Finally, the demo involves the search for a specific "target box", recognizing it amongst other boxes.

3 Technologies Inside the Demonstrator

The implementation of the use case required a great integration effort, bringing together and mixing different types of algorithms to produce a refined estimate of user pose and position.
The system was also intended as a test-bench to assess different computer vision algorithms and user interfaces to determine user satisfaction. We were able to highlight pro and cons for each methodology and to generate guidelines for further developments.

3.1 Mobile Hardware/Software Platforms

A Sony Xperia Z device [3] was selected as the principal hardware platform, together with the Sony Smartwatch as an alternative input-output device. All of the phases of the demonstrator have been implemented as Android applications, called by a main application. To be really useful for visually impaired users, all of the implemented algorithms were tested and customized to meet real-time requirements. Aural feedback methods were evaluated by the users, fed to them in the form of alerts, suggestions, instructions, primarily through the Text To Speech (TTS) Android engine. The Smartwatch was used to trigger the various phases of the demonstration, and the display on the phone was used as an explanation for guests at demonstrations, since all instructions to the user were acoustic or vibration.

3.2 Pedestrian Dead Reckoning

PDR was used to guide the user from the bus stop to the entrance of the office block. Although in the outdoor context, GPS could have been utilised, we wanted to demonstrate how a pedometer-based system could provide sub two-metre accuracy, as urban environments often deliver poor GPS accuracy. The PDR system is based on a combination of an inertial measurement unit, map-matching and advanced positioning algorithms. Building on the PDR approach proposed in [15], the proposed PDR uses MEMS and the CPU embedded in a Smartphone. The PDR solution includes walking and standing identification, step detection, stride length estimation, and a position calculation with a heading angular sensor. The localization of the pedestrian is estimated using the tri-axial accelerometer, gyro and magnetometers. Step detection and walking distance are obtained by accelerometers while the azimuth is computed by the fusion of gyro and map data. As is shown in Fig. 1, the PDR consists of 3 main modules: the pedometer module (PDR Engine), the Heading drift Compensation and the Map matching. The pedometer counts each step a person takes by detecting the vertical motion of the centre of gravity of the body, searching for patterns in the vertical acceleration which are matched to steps. Speed and stride length are associated to a step and are used by the PDR to compute new positions [22]. Since each person has a different walking physiological model, a calibration phase is needed to estimate the step length. The calibration stage

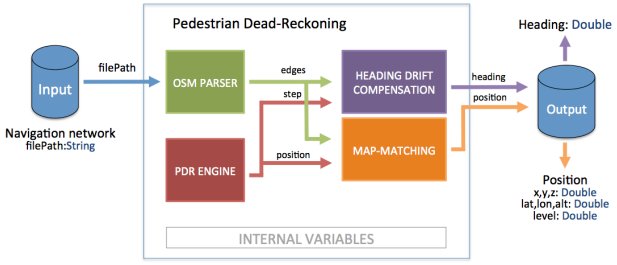

Fig. 1. Pedestrian Dead Reckoning block diagram

consists of a 30 meters walk. This calibration value is unique and reflects the walking characteristics of each pedestrian. Map data is also used to enhance the information provided by the PDR through a process known as map-aided positioning. Map-matching is one of the key elements to be used with a PDR module to obtain a more precise localization. Thanks to the structured indoor environment (straight lines defined by corridors and walls, with dominant directions), a user's trajectory can be matched to an edge described in the OpenStreetMap format, and a projected position can be determined. This map-matching is not widely used in PDR projects but it brings a better accuracy to CV algorithms. Because of numeric integration, the device direction computed from gyroscope attitude drifts, and hence a Heading Drift Compensation (HDC) algorithm [12], using map data, has been implemented. This algorithm uses straight-line features from the OpenStreetMap navigation network to correct, when possible, the heading derived from the gyroscope. At each detected step the HDC method finds the nearest straight-line from user's position and determines the offset to be applied to the heading.

3.3 Visual Navigator

The Visual Navigator is activated as soon as the user reaches the entrance of the building. This module coordinates one or more Visual Target Detectors and interacts with the visually impaired user to instruct her/him to safely reach a predefined destination. The interaction is managed by means of spoken messages (from the system to the user) and via a Smartwatch (bidirectional). The Visual Navigator receives, from the main application, a list of visual target detectors to be activated, in sequence or in parallel, in order to reach a set of local destinations. It notifies the user as soon as a target is detected in the scene, it communicates the direction to follow in order to reach it, and the action needed to pass from one target to the next. Moreover, the navigator warns the user in the case that dangerous situations are detected (i.e. a wet floor sign), vibrating the Smartwatch.
The Visual Target Detector module is capable of detecting and localizing a specified target within the visual range of the Smartphone camera. A target can be single- or multi-parts and its structure is encoded as a custom description

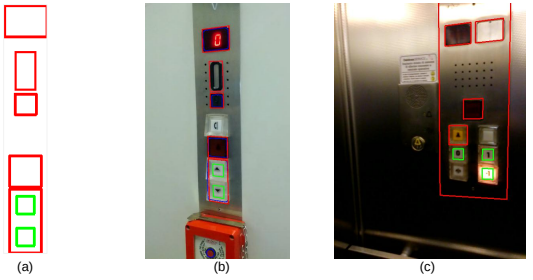

Fig. 2. (a) Geometric structure of a target, (b) Projection of the target structure on the image according to the estimated homography for the OutsideLiftButtons and (c) for the InsideLiftButtons detectors

file (xml format). The description file stores information about the number of sub-parts, their shape/size in the real world (represented by polylines) and the spatial relationships between them. The different parts are assumed to be co-planar. In Fig. 2a one can see a visual representation of the geometric structure of a multi-part target taken from an xml description. For each part, a specialized detection routine can also be specified in the xml file.
The Visual Target Detector works as follows:

- Specialized routines are applied to the input camera image to obtain a list of candidates for each part;
- Combinations of candidates are analysed in order to select the one most compatible with the target structure, taking into account the homography that maps the coordinates of the candidate regions to the real world one;
- The homography corresponding to the best combination is used to localize each part of the input image and to estimate the pose of the camera with respect to the target (using camera calibration data). The introduction of a multi-part target enables the detector to be robust to partial occlusions, a useful feature especially in the case of detecting lift buttons which can be covered by the user's hand.

Using fast template matching methods [27], text in scene algorithms [25] and skin detection techniques [20], we have developed nine detectors, seven of them are composed of a single part (WetFloor, FBK_panel, AR_Logo, LiftSign, ToyShop, LegoLogo, FingerTip) and two are composed of several parts (InsideLiftButtons, OutsideLiftButtons), see Fig. 3.

The Visual Navigator can activate Visual Target Detectors as desired. Initially in the demonstrator, it guides the user from the office entrance to the lift. Along the route, it checks for the presence of a wet floor sign (if found it warns the user). Once the first landmark (FBK_panel) is detected, the user is guided towards it by means of spoken messages with the purpose of maintaining

Fig. 3. Example of detectable targets

the target in the middle part of the camera's field of view. When the estimated distance between the user and the target falls below two meters the Visual Navigator launches the AR_logo detector. In a similar way, the system looks for the target and guides the user towards it. Again, when the distance is below 0.8 meters, the LiftSign detector is activated and the user is instructed to aim the camera in the correct direction in order to look for the lift sign. When the estimated distance of the target falls below 2 meters, the user is deemed to be close to the lift. The user is then invited to turn right to search for the button pad of the lift and the corresponding detector is activated (OutsideLiftButtons). When the user is close enough, she/he is invited to touch the button pad with a finger and the FingerTip detector is enabled. The system estimates the position of the fingertip with respect to the "arrow up" button and instructs the user how to move her/his finger, until finally the finger covers the correct button. The Visual Navigator invites the user to press the button and to wait for the lift. When the user enters the lift and taps the Smartwatch confirming to be inside, the Visual Navigator enables the InsideLiftButtons detector and the FingerTip detector is used to guide the user's finger to reach and press the "Floor 1" button. The Visual Navigator is again reactivated when the user is in the proximity of the toyshop. It then launches the ToyShop detector and guides the user towards the shop entrance. Finally, when the user is confirmed to be inside the shop, by tapping the Smartwatch, the LegoLogo detector is launched, and the navigator guides the user towards the Lego boxes shelf. A template matching algorithm, exploiting local features, allows the user to find the desired box.

3.4 Scene Classification

In the scenario of indoor navigation and shopping assistance, algorithms which exploit visual cues are fundamental for recognizing some contingent events, that cannot be derived only from the map of the building and the user's position.

For this reason, a visual scene classification algorithm has also been integrated into the system for detecting if a certain path is crowded, in order to perform a re-routing and help the visually impaired person to walk in a freer area, and also to detect if the user's device is facing an aisle or a shelf, inside the shop. The scene classification module can detect a set of different semantic classes, decided a priori during a training phase, extracting information from the images. Different problems should be considered in designing a scene recognition engine to be implemented on a mobile device for personal assistance: memory limitation, low computational power and very low latency, to achieve real-time recognition. In a recent work [18], the DCT-GIST image representation model has been introduced to summarize the context of the scene. Images are holistically represented starting from the statistics collected in the Discrete Cosine Transform (DCT) domain. Since the DCT coefficients are already available within the digital signal processor for the JPEG conversion, the proposed solution obtains an instant and "free of charge" image signature. This novel image representation considers the Laplacian shape of the DCT coefficient distributions, as demonstrated in [23], and summarizes the context of the scene by the scale parameters. The recognition results, obtained by classifying this image representation with Support Vector Machines [14], closely match state-of-the-art methods (such as GIST [26]) in terms of recognition accuracy, but the complexity of the scene recognition system is greatly reduced. For these reasons, this approach has been specialized for the classes needed by the defined use case. More specifically, the indoor navigation scenario could take advantage from the recognition of 4 different categories:

1. Crowded/Non-crowded;
2. Aisle/Shelf

A database has been built downloading images from the Internet and chosing about 600 samples per class; images have been chosen to represent real conditions, in terms of variety of illumination, location, etc.; then the database has been used to train 2 SVM classifiers for detecting each pair of categories. A cross-validation approach was used for assessing the performances of the algorithm on these classes. In particular, ten training and testing phases were executed for each pair of classes: at each iteration 90% of the database is used for training and the remaining 10% for testing. Finally, the global accuracy is obtained by averaging the accuracies of the ten cross-validation phases. Results are reported in Table 1.

It is quite evident that the DCT-GIST scene representation has very good performance on classes required by the indoor navigation and personal assistance system, and it is suitable for real-time applications. This algorithm was

Table 1. Scene classification results on the categories needed by the use case

	Crowded	Non-Crowded
Crowded	0.95	0.05
Non-Crowded	0.04	0.96

	Aisle	Shelf
Aisle	0.93	0.07
Shelf	0.07	0.93

integrated with the heading information from the pedometer to provide the user with instructions for turning to her/his right and left to explore which corridor is less crowded; then the freer passage is chosen to continue the navigation.

3.5 Structure from Motion

Navigation from the lift exit to the shop entrance is achieved by a SfM algorithm, integrated with some information from the PDR. As shown in Fig. 4, the implemented SfM pipeline consists of two main blocks. The first one processes the incoming images, in order to identify salient elements in a scene, like corners, and tracks them in the sequence to obtain a sparse optical flow. We compute a set of features per frame and generate the Optical Flow (OF) using descriptors (i.e. a vector of coefficients, able to characterize the pixels around a corner) which are compared during the matching phase, thanks to a specific function to verify an association's correctness [13]. The second block, also called the Geometry stage, uses geometric relations (like the epipolar constraint), to estimate the camera position and orientation by processing the OF produced in the previous phase. This strategy processes 2D-2D associations, finally computing a fundamental matrix thanks to a family of algorithms called N-points algorithms, where "N" indicates the number of random points engaged in the estimation of a camera pose. To obtain a robust solution, RANSAC [19] is combined with an N-points algorithm, and then the corners are triangulated to obtain a sparse 3D representation of the environment (3D Map).

A fundamental assumption, for a successful reconstruction, is the knowledge of internal camera parameters. The procedure is achieved by estimating the camera's focal length, the centre of the image and the lens distortion through the prior acquisition of a set of images framing a checkerboard. This functionality has been integrated inside the application in order to easily recalibrate the system, each time the demo is installed on a new Android device.

The SfM reconstructions are scaled in order to fit the reconstruction into the real world, this is done by fusing visual information to inertial sensor data. In this way, the SfM trajectory is rescaled using the PDR pedometer step size information, discussed in Section 3.2. This real-time rescaling enables a conversion of the user's position from euclidean space to geographical space.

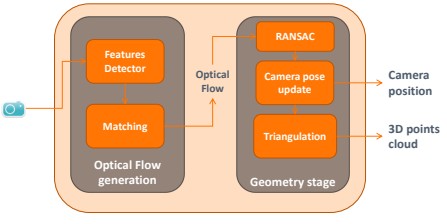

Fig. 4. Pipeline of the Structure from Motion approach

4 User Experience Evaluation

Beyond the indoor navigation and augmented reality experience, it is clear that future assistive applications will also need new interaction and application models to facilitate new forms of communication and meet increasingly high user expectations [11]. This is a huge challenge since assistive technologies for visually impaired people cannot rely on design guidelines for traditional user interfaces. New user interfaces permit interaction techniques that are often very different from standard WIMP (Windows, Icons, Menus, Pointer) based user interfaces [17]. Digital augmentation, in general, can include different senses such as sight, hearing and touch, hence they must be realized with an array of different types of input and output modalities such as gestures, eye-tracking and speech. Aspects such as: finding out how well a user performs tasks in different contexts; letting the user interact with natural interfaces; and hiding the complexity of the technology; are central to ensure good quality applications and a good user experience. In the light of these considerations, during the development of the navigator, user studies in indoor and outdoor scenario were conducted to assess social acceptance and to receive feedback for further developments. Aural guidance effectiveness was evaluated, conducting preliminary tests with blind and visually impaired people to receive proper feedback, as described in Sec. 4.1. A visualization was also created and evaluated, prior to the completion of the demonstrator, as described in Sec. 4.2. The received feedback drove the development in the last part of the implementation providing the improvement of the audio interface and the rules the device should use in these types of applications. The last evaluation, performed by experts in the field, was done when the entire system was integrated (Sec. 4.3). Ideally, the system should have been evaluated with a user study of visually impaired people. This was unfortunately not possible for all of the phases, but only for the outdoor navigation part, but it is recommended as future work to fully understand the improvements needed to fulfil the needs of the intended user group.

4.1 Evaluation of the Aural Guidance

During the development of the system, an analysis of the aural guidance technology was conducted, with several tests in the field:

1. March, 2013: Indoor and outdoor tests in Grenoble, with Alain De Borniol who is visually impaired (president of the association ACP: Accès Cible Production) and Thomas Marmol (blind son of Bruno Marmol, IT manager at INRIA);
2. April, 2013: Indoor tests at INRIA Rhône-Alpes with Christine Perey, sighted and chair of the AR Standards Community;
3. June, 2013: Indoor tests in Sugimotocho railway station and outdoor tests in Osaka city with visually impaired students from the school for the blind.

After these studies, feedback was collected:

1. Let the user choose their favourite text-to-speech engine for the navigation application;
2. Audio instructions are too long;
3. Prefer right, left, in-front or behind terms instead of using indications in degrees;
4. Make the application compliant with Google Talkback [4];
5. Interactions with the Smartphone are difficult during navigation; when possible a remote controller would be appreciated;
6. Instructions must be repeatable.

According to this feedback, some corrective actions have been implemented: first of all the application has been modified to take into account the chosen TTS engine in the general Android settings, instead of the default one (Pico TTS); guidance instructions have been focused on the nearby environment (5 meters for indoor and 20 meters for outdoor); audio instructions were changed, in accordance to feedback 3 and compliance with Talkback was guaranteed. To satisfy requirement 5, the application was modified to interact with different external devices, such as a headset, an ear set and a watch; further tests should be conducted to decide which external devices are the most suitable. Finally, to solve issue 6, the application was modified to permit the user to ask for the latest instruction by a simple pressure on the watch, the headset or the ear set.

4.2 Evaluation on the Visualization

The implemented use case consists of multiple phases. To avoid the fragmentation of the whole demonstrator, it was decided to try to visualize it before its final integration. Visualization has the objective to elicit feedback in areas that potentially require additional development. It was decided to simulate the personal assistance demonstrator using a Wizard of Oz (WOZ) approach. The WOZ prototyping method is particularly useful to explore user interfaces for pervasive, ubiquitous, or AR systems that combine complex sensing and intelligent control logic and is widely used in human-computer interaction research [16]. The idea behind this method is that a second person can simulate the missing parts of a system. Hence, it was decided to use the WOZ tool called WozARd [8], [5], to simulate the system and do a pre-evaluation prior to development completion. All phases (see Sec. 2.2) were setup at Lund University, Sweden. Two Xperia Z phones were used; one by the test leader (Wizard) and one was attached to a backpack, carried by the subject. The subject wore a Smartwatch to interact with the application. A video showing the simulation of the demonstrator can be found in [6]. The subject's eyes were covered and everything was recorded with a video camera. At the end of the experiment, suggestions and issues emerged from the video. The first suggestion was to start an application with a long press on the Smartwatch to activate a "voice menu" saying "Lego shopping", "walking home", etc.. The start of the Tour could also be initiated through scanning a NFC tag. The user should be allowed to skip assistance for a certain phase (but hazard detecting remain running), because he could be bored by excessive indications. When a new phase is started, feedback or summary might be needed;

the information should tell the user what will come next from that point to the next phase. In front of the elevator, it would be better not to guide the hand but instead tell the user "you can now call the elevator" when her/his hand is on the correct button. The guidance to the correct product, requires the user to take the phone and use it to search the shelves. When the product is found, maybe it would be good to allow for product ID confirmation through a NFC or barcode scan as well as have user confirmation that she/he has completed the product navigation part. Moreover, a general improvement would be gained by a careful design interaction paradigm for watch, phone and voice input: each command on the watch (e.g. double tap, long press, swipe left, etc.) could activate a different action, but it is fundamental to avoid accidental selections.

4.3 Expert Evaluation

The VeDi system has been developed taking the recommendations in Sections 4.1 and 4.2 into account. The hardware/software platform was presented and demonstrated for expert evaluation in December 2013. The demonstrator was adapted for that location, hence the outdoor and indoor navigation would need modifications to replicate it in other locations. The evaluation was conducted through invited experts, from the Haptimap EU project [7], with extensive experience in navigation systems and with visually impaired people. As is commonly done in user experience reviews, the following feedback concentrates solely on issues found during the demonstration, for deriving suggestions to improve the system.

1. **Tutorial/help** The current version of the system does not have any manual or tutorial. Without an extensive tutorial or manual it is unlikely that a user that has never used the system can use it. To learn the system before using it would be important since live usage will contain situations where user misunderstandings may lead to dangerous or at least socially awkward situations, which may result in low interest in using the system again.
2. **System latency** Some parts require the system to perform heavy computations in order to deliver instruction to the user. In these situations the instructions arrive a little late. During this delay the user moves and makes the next instruction incorrect. Each part of the system should be tested and delays measured. Viable solutions may vary depending on situations and delays. One solution could be to make the user aware of system latency and advise her/him when to move slowly. Another solution could be to change the duration and type of feedback to minimize the delay effect. In the worst case, some delays render the assistance unacceptable and the system must improve to become usable.
3. **Camera orientation** Some algorithms need the Smartphone to be in landscape mode, whilst others require portrait. The system was designed to be worn in a chest harness so that the user has both hands available for other tasks. It is highly recommended that all algorithms can function in portrait mode to avoid the need to switch orientation.

4. **Amount and type of instructions** The system has many audio-based instructions and feedback sounds. Visually impaired people use their hearing to compensate for their lack of full vision. For these people, audio may cause a high cognitive workload and many visually impaired people react negatively to using heavy audio solutions. It is recommended to keep the duration and amount of audio instructions to a minimum yet on an understandable level, a clear design challenge. Users with different experience levels of the system or in different locations may require different amounts of instructions. For example, if a routine has been carried out daily for months, a user may not require the same number of confirmations that a new user would need. It is recommended that the system have different levels of verbosity and a way for the user to control it.
5. **Assistance versus instructions** The goal of a navigation assistant for visually impaired people is to give autonomy and personal freedom to the person. If this is the case, should this autonomy also be from the assistant itself? It is the experience of the reviewers that the assistance desired by the visually impaired is a case of personal preference. It is recommended that the system enables users to choose alternative paths or activities during assistance. You could think of the system as a store clerk that you can ask for recommendations but the decision is still up to the user.

5 Conclusions

A hardware/software system for guiding a visually impaired person through a building to a store shelf to find a specific item was presented. We showed an integration of vision-based with pedestrian localization systems and how we created an assistant for indoor/outdoor navigation. Building this system required a large integration effort to merge all the vision-based algorithms with hardware-sensor pedometer, to derive a constant estimate of the user's position. Algorithms were also chosen and customized to meet real-time execution requirements, which is crucial for assistive technologies on mobile devices. Many user studies were conducted during the project evolution and corrective actions were applied to the final demonstrator. Moreover, an expert review of the final demonstrator resulted in a number of recommendations that could be used to significantly improve the usability of the system. Our research has confirmed that building a technology for the assistance of the visually impaired requires a deep user study to iteratively assess user satisfaction and then to bring improvements and corrections to the system accordingly.

Acknowledgements. This research is being funded by the European 7th Framework Program, under grant VENTURI (FP7-288238). The scene classification part has been developed within the Joint Lab between IPLAB and STMicroelectronics Catania.

References

1. WHO, Fact Sheet N°282: "http://www.who.int/mediacentre/factsheets/fs282/en/"
2. OpenStreetMap official website: "http://openstreetmap.org"
3. Sony Xperia Z: "http://www.sonymobile.com/global-en/products/phones/xperia-z/"
4. Google Talkback: "https://play.google.com/store/apps/details?id=com.google.android.marvin.talkback"
5. WozARd: "http://youtu.be/bpSL0tLMy3w"
6. Simulation of VENTURI Y2D: "http://youtu.be/NNabKQIXiTc"
7. HaptiMap project, FP7-ICT-224675: "http://www.haptimap.org/"
8. Alce, G., Hermodsson, K., Wallergård, M.: WozARd. In: Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services - MobileHCI. pp. 600–605 (2013)
9. Angermann, M., Frassl, M., Doniec, M., Julian, B.J., Robertson, P.: Characterization of the Indoor Magnetic Field for Applications in Localization and Mapping. In: 2012 International Conference on Indoor Positioning and Indoor Navigation. pp. 1–9 (2012)
10. Baniukevic, A., Jensen, C.S., Hua, L.: Hybrid Indoor Positioning with Wi-Fi and Bluetooth: Architecture and Performance. In: 14th International Conference on Mobile Data Management. vol. 1, pp. 207–216 (2013)
11. Barba, E., MacIntyre, B., Mynatt, E.D.: Here we are! Where are we? Locating mixed reality in the age of the smartphone. In: Proceedings of the IEEE. vol. 100, pp. 929–936 (2012)
12. Borenstein, J., Ojeda, L., Kwanmuang, S.: Heuristic Reduction of Gyro Drift in a Personal Dead-reckoning System. Journal of Navigation 62(1), 41–58 (2009)
13. Brox, T., Bregler, C., Matas, J.: Large displacement optical flow. In: IEEE Conference on Computer Vision and Pattern Recognition. pp. 41–48 (2009)
14. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology 2(3), 27:1–27:27 (2011)
15. Colbrant, A., Lasorsa, Y., Lemordant, J., Liodenot, D., Razafimahazo, M.: One Idea and Three Concepts for Indoor-Outdoor Navigation (2011), INRIA Research Report n° 7849
16. Dow, S., Macintyre, B., Lee, J., Oezbek, C., Bolter, J.D., Gandy, M.: Wizard of Oz Support throughout an Iterative Design Process. IEEE Pervasive Computing 4(4), 18–26 (2005)
17. Dunser, A., Billinghurst, M.: Handbook of Augmented Reality, chap. 13, pp. 289–307. B. Furht (Ed.) (2011)
18. Farinella, G.M., Ravì, D., Tomaselli, V., Guarnera, M., Battiato, S.: Representing Scenes for Real-Time Context Classification on Mobile Devices. "http://dx.doi.org/10.1016/j.patcog.2014.05.014" (2014)
19. Fischler, M.A., Bolles, R.C.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Communications of the ACM 24, 381–395 (1981)
20. Jones, M.J., Rehg, J.M.: Statistical color models with application to skin detection. International Journal of Computer Vision 46(1), 81–96 (2002)
21. Kurata, T., Kourogi, M., Ishikawa, T., Kameda, Y., Aoki, K., Ishikawa, J.: Indoor-Outdoor Navigation System for Visually-Impaired Pedestrians: Preliminary Evaluation of Position Measurement and Obstacle Display. In: ISWC '11 Proceedings of the 2011 15th Annual International Symposium on Wearable Computer. pp. 123–124 (2011)

22. Ladetto, Q.: Capteurs et Algorithmes pour la Localisation Autonome en Mode Pédestre (2003), phd Thesis, École Polytechnique Fédérale de Lausanne
23. Lam, E., Goodman, J.W.: A mathematical analysis of the DCT coefficient distributions for images. IEEE Transactions on Image Processing 9(10), 1661–1666 (2000)
24. Le, M.H.V., Saragas, D., Webb, N.: Indoor navigation system for handheld devices (2009), master's thesis, Worcester Polytechnic Institute, Massachusetts, USA
25. Messelodi, S., Modena, C.M.: Scene Text Recognition and Tracking to Identify Athletes in Sport Videos. Multimedia Tools and Applications 63(2), 521–545 (2013)
26. Oliva, A., Torralba, A.: Modeling the shape of the scene: a holistic representation of the spatial envelope. International Journal of Computer Vision 42(3), 145–175 (2001)
27. Porzi, L., Messelodi, S., Modena, C.M., Ricci, E.: A Smart Watch-based Gesture Recognition System for Assisting People with Visual Impairments. In: ACM International Workshop on Interactive Multimedia on Mobile and Portable Devices. pp. 19–24 (2013)
28. Xiao, W., Ni, W., Toh, Y.K.: Integrated Wi-Fi fingerprinting and inertial sensing for indoor positioning. In: 2011 International Conference on Indoor Positioning and Indoor Navigation. pp. 1–6 (2011)