ARIANNA: a smartphone-based navigation system with human in the loop

Daniele CROCE, Pierluigi GALLO, Domenico GARLISI,
Laura GIARRÉ, Stefano MANGIONE, Ilenia TINNIRELLO
DEIM, viale delle Scienze building 9, Universitá di Palermo, 90128 Palermo - Italy
Email: <name>.<SURNAME>@unipa.it

Abstract— **In this paper we present a low cost navigation system, called ARIANNA, primarily designed for visually impaired people. ARIANNA (pAth Recognition for Indoor Assisted NavigatioN with Augmented perception) permits to find some points of interests in an indoor environment by following a path painted or sticked on the floor. The path is detected by the camera of the smartphone which also generates a vibration signal providing a feedback to the user for correcting his/her direction. Some special landmarks can be deployed along the path for coding additional information detectable by the camera.**

In order to study the practical feasibility of the ARIANNA system for human users that want to follow a pre-defined path (by only using the smartphone feedback signals), we study how to incorporate human behavior models into the feedback control loop. We also implement an Extended Kalman Filter for localization, in which the user coordinates, speed and orientation represent the filter state (whose updating law depends on the user reaction to the vibration signals), while the smartphones sensors provide the set of measurements used for state estimation.

Index Terms— **navigation system; vibration; assistive technology; kalman filtering;**

I. INTRODUCTION

Outdoor navigation based on GPS signals is a common technology which is nowadays included in many off-the-shelf smartphones. Many GPS-based applications are available in the market for aviation, naval and terrestrial uses. Recently the research community has focused the attention on indoor navigation, where GPS is not available. Navigation requires target localization, which can be done using different methods, such as triangulation of RF signals (mainly WiFi), direct sensing (with RFIDs, ultrasound, bluetooth, etc.), pattern matching, dead reckoning based on odometry readings (accelerometers, magnetometers, compasses, and gyroscopes). For example, dead-reckoning techniques are employed in Navatar [1] where users interact with the application and help correcting possible navigation errors. RF-PATH-ID [2], instead, is based on disseminating passive RFID tags and using a dedicated reader to acquire information on the user location. More examples and detailed information on indoor localization techniques may be found in [3].

Assistive tools for indoor navigation have specific requirements in terms of reaction time (they require to run in real-time to be useful) so they need an adequate refresh frequency. Tools must be light-weight, portable, low-power and low-cost and should require minimum training time. Solutions based on off-the-shelf devices can be easily spread, even better if the used devices are already available to people. In this sense, smartphones are light-weight, portable, and affordable devices that are already in everyone's pocket. For this reason, we exploit computer vision capabilities of common smartphones provided with cameras and present "path following people" accordingly to some key ideas presented in [4]. In the present paper, ARIANNA (pAth Recognition for Indoor Assisted NavigatioN with Augmented perception) is described, equipped with a robust tracking system based on an Extended Kalman Filter (EKF) that estimates the states from noisy observations. Kalman Filters (KF) are widely used in computer vision and robotic systems for object tracking, path following, simultaneous localization and mapping (SLAM), leader-follower systems and 3-D modeling, just to cite a few. Many applications of the KF in robot vision are summarized in [5]. With the help of the EKF-based tracking, the proposed system recovers even in case the path is temporary lost by providing a shift to a new user-centric perspective, where the navigating user runs corrective actions and is a controller in the interactive control system. The human intervention in interactive control systems is named 'human-in-the-loop' (HIL), [6], [7]. A key part of the HIL control are human responses to stimuli: they depend on physiological, psychological and environmental factors. For example, several alternative paths might be taken due to unexpected obstacles, orientation disorders, sleepy conditions, different step lengths, etc. Primarily designed for visually impaired people, the ARIANNA system is a particular example of HIL feedback control system.

Regarding the specific case of visually impared, many recent technologies have been developed to help them move autonomously in unfamiliar environments and different interfaces have been designed to communicate with the visually impaired. For example, virtual acoustic displays and verbal commands issued by a synthetic speech display are used in [8]. AudioGPS [9] and Melodious Walkabout [10] use audio cues to provide information on the surrounding environment. However, acoustic feedback is perceived as a distraction and overloads visually impaired hearing which is already used to catch information on the near environment. It is thus preferable to avoid audio indications in favor of tactile alternatives. Indeed, haptic principles and a list of possible applications are presented in [11], while benchmark metrics for haptic interfaces have been recently proposed, based on a combination of physical and psychophysical data [12]. Frictional forces arising from the stroke of a finger

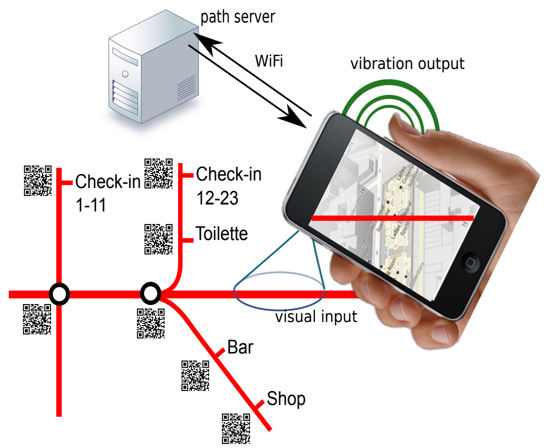

Fig. 1. ARIANNA navigation system description

moving on a surface are studied in [13], [14], while tangential skin displacement at the fingertip (stimulus speed and displacement length) communicate direction or displaying static friction in haptic applications [15]. In [16], vibrational feed-back is given by a special glove in the Finger-Braille language. This system requires some dedicated hardware and is specific to the language used. Other examples of haptic feedback can be found in [17], [18]. Compared to other solutions, the approach we will present hereafter is much easier to use because it employs the smartphone as a visual-to-vibration translator for directional information.

II. Navigation system description

The ARIANNA navigation system is especially designed for indoor scenarios, where GPS-based solutions are unavailable, and exploits the visual sensor and vibration signals of commodity smartphones. Figure 1 shows a typical use of ARIANNA in an airport scenario. The paths of interest are marked with colored lines on the floor. This is the only dedicated instrumentation applied to the environment and is a quite simple and low cost solution; QRcodes are settled close to points of interest, on line intersections and are also used for landmarking. They provide information on the right line to follow in order to get to the desired destination. The user interface employs tactile stimuli to receive feedback on the heading corrections to be employed, as better described in the following. The systems itself is composed by five main components: (A) ambient instrumentation; (B) sensors; (C) data transport network; (D) path server; (E) tactile interface.

A. Ambient instrumentation

The ambient instrumentation is composed of colored tapes which can be easily sticked on the floor or carpets to define different paths. This is the only dedicated instrumentation applied to the environment. Paths can intersect each other forming a planar graph where intersections are nodes of the graph. To add information on the paths, any segment (the graph edges) may be deployed with two parallel strips with different colors, so the ordered couples (color1, color2) and (color2, color1) encode both direction and orientation. Additionally, using bar codes or QRcodes it is possible to encode relevant information regarding the edges (as for example the distance from/to the extremes of the segment) and for landmarking.

B. Sensors

The main sensor used in the ARIANNA system is the camera, which most smartphones on the marketplace are equipped of. The camera is used to reveal the presence of lanes on the floor and acts as a visual control for the haptic transducer. We also use the embedded compass and accelerometer sensors to help maintain or recover the visual control of the line in the EKF-based tracking. All these sensors are available on most commodity smartphones: this is a key aspect for keeping the system low-cost and affordable for a vast public.

C. Data transport network

We assume that a data network connection is available for downloading the ambient map (e.g. via a WiFi or 3G network). The data transport network does not require specific adaptations but is a facility that permits communication between the phone and the ARIANNA server. The server is used to provide localization information, correlation between paths and points of interest, routing towards the destination. The presence of the server and the wireless network is necessary only in case the application is unaware about the building topology and its deployed paths. On the contrary, if the application loaded on the phone has such information locally available, the presence of network and server is optional (even if flexibility is possible only with those elements, as explained below).

D. The path server

The path server stores and retrieves information from a path repository via the url printed into the QRcode. The content pointed out by the (fixed) url can be changed on the fly with a simple update on the server. Such flexibility permits path adaptation required by topological changes due to maintenance or load balancing. When the smartphone detects a QRcode on the path, it immediately runs an http request to the server using the url inside the QRcode. The server knows the position of the user (because of its proximity to the QRcode position) and sends back to the smartphone the next edge to follow. In fact, among all paths deployed in the building, thanks to the indications provided by the path server, the smartphones provides haptic feedback only towards the "enabled" paths according to the server indication.

E. Tactile Interface

The tactile interface is a key point of the system. The behavior of the haptic feedback can be summarized as follows: the camera continuously grabs the scene in front of the person. The tracking system incorporates the information on the line (together with the compass and accelerometer data) and provides feedback with the phone vibration. The intensity and type of the vibration is based on the output of

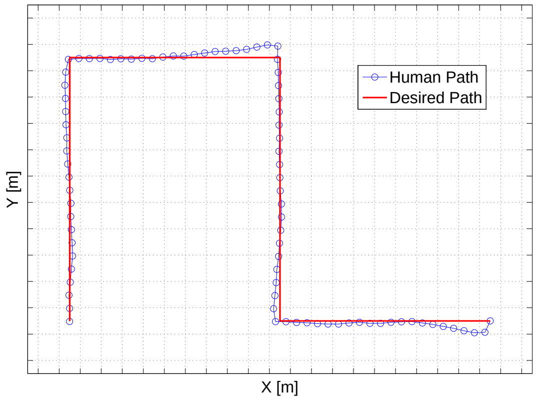

Fig. 2. Human path with wrong initial orientation, $m = 0.9$ and $n = 0.4$.

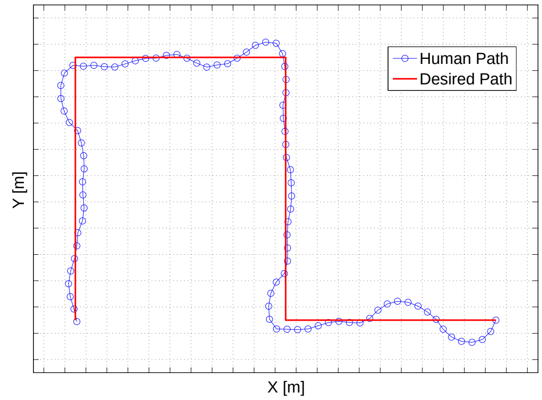

Fig. 3. Human path with wrong initial orientation, $m = 0.6$ and $n = 0.4$.

the EKF and is designed to keep the camera always in contact with the line or to bring back the visual contact when it is lost. Vibration is a native functionality of the phone obtained through a rotating eccentric mass. It has been shown that the current consumption of typical vibration motors has a limited impact on the battery life of commercial smartphones[19] and that the energy savings coming from switching off the screen are higher than the costs introduced by vibrational cues [20]. Unlike other approaches in haptic interfaces, our solution does not need a selective vibration of the touched point (that is also difficult to obtain and requires special piezo-electric materials, etc.).

III. Human in the Loop

In order to study the practical feasibility of the ARIANNA system for human users that want to follow a pre-defined path (by only using the smartphone feedback signals), we need to determine how to incorporate human behavior models into the formal methodology of feedback control as in [6].

For sake of simplicity, we assume that smartphone and user positions coincide and movements are possible in a 2D environment (i.e. we do not consider changes in user elevation). The paths are represented by a piecewise constant function of the space that is different from zero along the path points. We also assume that the user starts navigating from a point belonging to the path and that the smartphone is able to observe a squared region of the floor whose dimensions are about 50 cm each. The goal of the application is allowing the user to reach the end point of the path. Since smartphone signals are generated and updated at discrete time steps, we consider a discrete time system in which the walking behavior of the user is updated at the same temporal scale of the feedback signals. Finally, we consider that the pedestrian speed is approximatively constant for a given user (although a real estimation of the speed is possible and will be discussed in the next section).

Let β be a generic constant direction of a portion of the path. User reaction to the smartphone signals is described in terms of a change of his heading direction α. Such a change is based on two different information provided by the phone: i) the user direction relative to the path $\beta - \alpha$, that is perceived according to the alignment of the vibration points on the phone display; ii) the distance between the user and the path, that is perceived according to the distance of the vibration points from the center of the phone display. It is reasonable to assume that the user will correct the walking direction by trying to be aligned to the path and to null the distance from the path in the next steps. Let α_k be the user direction at the discrete time k, v the user speed, and T_X the desired maximum time interval for nulling the path distance.

We consider the following *human in the loop* model. When the smartphone is able to see the painted line on the floor, the user heading direction α_{k+1} is updated by considering a first correction proportional to the perceived deviation from β (i.e. $\beta - \alpha_k$), and a second correction proportional to the direction required for nulling the path distance by T_X (i.e. $sin^{-1}\frac{d_k}{v \cdot T_X}$). When the path is not captured by the phone camera, the user heading direction is corrected by an angle with constant module Δ, whose sign is positive (negative) if the path was lost on the left (right) side of the user. Being T the discrete time step of feedback and movement updates, we have:

$$\alpha_{k+1} = \begin{cases} \alpha_k + m(\beta - \alpha_k)^u + n\sin^{-1}\frac{d_k^u}{v \cdot T_X} & |d_k| \leq 50cm \\ \alpha_k + sign(d_k) \cdot \Delta \quad |d_k| > 50cm \end{cases}$$
$$d_{k+1} = d_k + v \cdot T \sin(\beta - \alpha_k)$$

where $(\beta - \alpha_k)^u$ and d_k^u represent the *human perception* of the signals $\beta - \alpha_k$ and d_k displayed at time k (that can be assumed equal to the real values plus an additive noise), and the coefficients m and n model the *human behavior*. Perception noises are generally assumed with zero mean, although a bias can be considered for taking into account the asymmetrical space perception of some users.

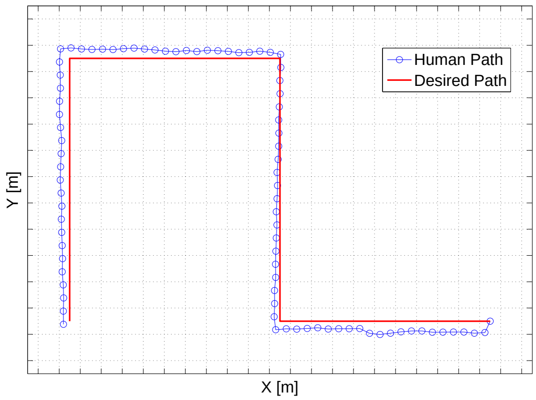

Fig. 4. Human path when distance corrections are almost neglected ($m = 0.9$, $n = 0.01$).

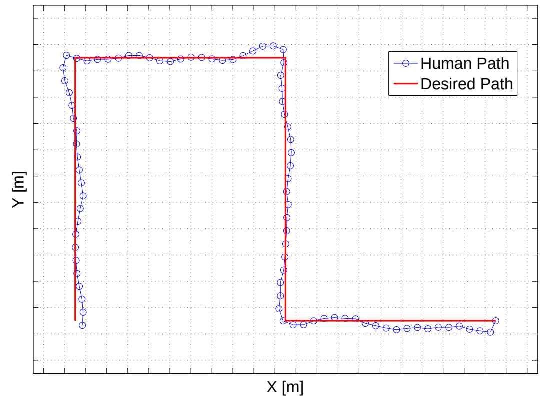

Fig. 5. Human path with an perception direction noise uniformly distributed in $[-\pi/6, \pi/6]$.

A. Examples of human behaviors

For visualizing the effects of different human perception and behavior models, we run some simulations in which at each time instant T the user coordinates are updated according to our human-in-the-loop model. All the simulations refer to the same path (the red lines plotted in figures 2-4) and have been obtained by setting $v = 0.5m/s$, $T = 1s$, $\Delta = \pi/10$. Perception noises and correction factors m and n have been used as configuration parameters for modeling different users.

Figure 2 shows the typical *normal* behavior, when the user starts its navigation from a point belonging to the path (the rightmost point of the figure) with a wrong orientation. Perception noise on the path direction has been assumed to be uniform in the range $[-\pi/15, \pi/15]$, while an additive Gaussian noise with zero mean and standard deviation equal to 5cm has been added to the perceived distance from the path. In this experiment, we set $m = 0.9$ and $n = 0.4$. Indeed, when the path is not lost (i.e. α_{k+1} is updated according to the $d_k \leq 50cm$ equation), we can easily study the system stability by considering that $\sin^{-1} d_k/(v \cdot T_x)$ can be linearized to $d_k/(v \cdot T_x)$. A good control design can be obtained if n is equal to about $m^2/(4T_x v)$. The figure clearly shows that after the transient phases to the wrong orientation (occurring at the beginning of the experiment and after each direction change), the user movements are almost on the path.

Since user real movements depend on the user strategy to follow the path and cannot be configured according to stability considerations, figure 3 plots the results of an experiment with non-optimal settings ($m = 0.6$ and $n = 0.4$). We can observe that the user is still able to reach the end of the path in an higher number of steps. In some cases, the user loses the path because the distance from the path is higher than 50cm. However, thanks to the second equation of the heading control ($d_k > 50cm$), the user is able to go back to the painted line.

Figure 4 shows another example in which corrections due to the distance from the path are almost neglected. In this case, the user reaches the end of the path by keeping an almost constant distance from the painted line. The distance is lower than 50cm, thus allowing to have continuous perception feedback. Finally, figure 5 shows a last example of human movements in presence of significant errors in the direction perception.

IV. TRACKING SYSTEM

Although our application has been designed for allowing users to follow a pre-defined path, we also envisioned the possibility to track the position of the users. To this purpose, we exploit not only the vision-based signals captured by the smartphone camera, but also a set of additional measurements provided by most smartphone models. The idea is calibrating or resetting the estimator of the user coordinates when the smartphone detects a reference point and integrating the information provided by the camera, compass, accelerometer, step counter, and so on, for updating the estimates when reference points are not available. The estimator is based on an Extended Kalman Filter in which we also include the user reaction to the phone signals in terms of an additive external signal on the state. Figure 7 shows the overall picture of the control system: on the basis of a state model describing the user movements and his reactions to the feedback signals, the measurements collected by the smartphone are filtered for producing an estimate of the user position and for helping in finding different destinations. When the user looses the path, vibration signals will drive him along a circular trajectory that permits to find again the instrumented path.

A. State Model

The paths are deployed on the floor as colored tapes, along which landmarks (e.g. QR codes) can be periodically applied for providing the absolute coordinates of the corresponding application point.

Let x and y be the 2D coordinates of the user, and $|v|$ and α the module and the direction of user velocity. We

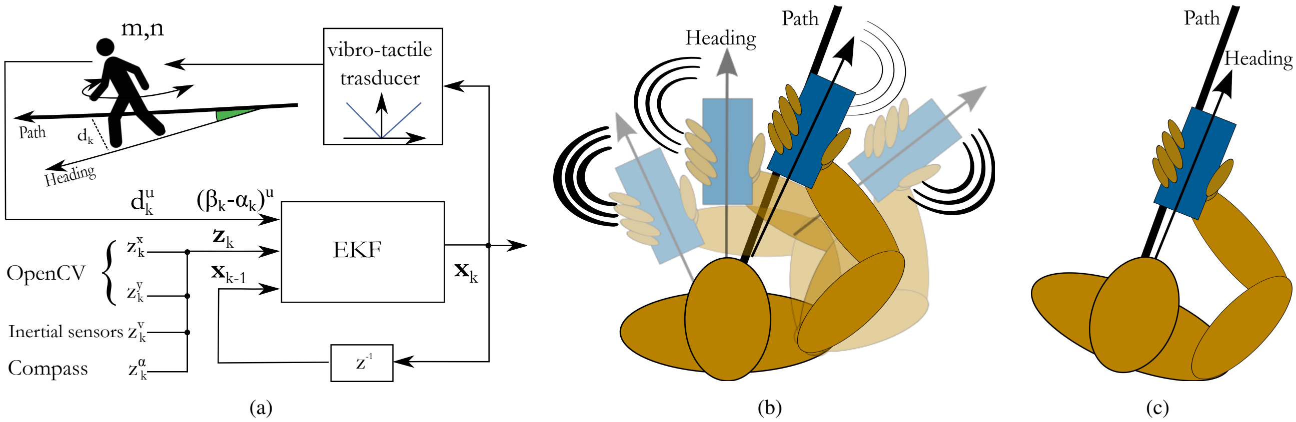

Fig. 6. The block diagram of the tracking system with an Extended Kalman Filter (EKF) (a); exploring and detecting the best heading direction through arm movements (b); rotation of the body to follow the desired direction (c).

chose to model the velocity in terms of module and direction (rather than in terms of orthogonal components v_x and v_y) because in the previous section we assumed that the human correction actions work on the direction of the movement. Being the user reaction signal u_k at time k equal to $m(\beta - \alpha_k)^u + n \sin^{-1} d_k^u/(v \cdot T_X)$ when the path line is visible to the phone and Δ when the path line is not visible, we can consider the following discrete-time state model:

$$\begin{cases} x_{k+1} &= x_k + |v|_k \cos(\alpha_k) T \\ y_{k+1} &= y_k + |v|_k \sin(\alpha_k) T \\ |v|_{k+1} &= |v|_k + w_{k+1}^{|v|} \\ \alpha_{k+1} &= \alpha_k + u_k + w_{k+1}^{\alpha} \end{cases} \tag{1}$$

where T is the update time interval, and $w_{k+1}^{|v|}$ and w_{k+1}^{α} are the state noise components. The additive noise on the velocity component represents the random variations in the pedestrian velocity that can be assumed as Gaussian distributed, while the noise on the user direction is given by the random fluctuations due to the real walking behavior of the user.

B. Measurement Model

The measurement model is based on the sensors available in the smartphone and on the information that can be inferred by the environment. Measurements are generally provided at regular time intervals T, but some specific components can be available only in some conditions (e.g. when the smartphone camera is able to read a landmark). More into details, we exploit the following measurements:

- *velocity (z^v)* : the user velocity is measured by some smartphone applications by exploiting the intertial sensors for counting the number of steps during an observation interval that can be assumed as an integer multiple of T. Such a mechanism requires to be calibrated to the user-specific step length. The measurement noise depends on the step detection sensors and on the approximation of fixed step length.
- *user heading (z^α)* : the direction of user movements is measured by means of the digital compass, which evaluates the user direction by referring to the south-north direction. The noise affecting this measurements is basically the compass noise. Additionally, computer vision techniques, based on optical flow concepts, can be used for providing another measurement of the heading direction (as well as another measurement of the user velocity module).
- *user coordinates (z^x and z^y)* : the user position can be read in the landmarks captured by the smartphone camera when they are visible in the current measuring interval T. This measurement is not always available and is affected by a noise representing the difference between the user coordinates at the end of the T interval and the real landmark position.

Being ν the generic noise on the measurement components, we can relate the measurements to the state vector by using the following equations:

$$\begin{cases} z_{k+1}^x &= x_{k+1} + \nu_{k+1}^x \\ z_{k+1}^y &= y_{k+1} + \nu_{k+1}^y \\ z_{k+1}^v &= |v|_{k+1} + \nu_{k+1}^{|v|} \\ z_{k+1}^\alpha &= \alpha_{k+1} + \nu_{k+1}^\alpha \end{cases} \tag{2}$$

where the apix of each noise component z has been explicitly related to the physical meaning of the relevant measurement.

C. Tracking Example

Figure 7 shows the results of three different experiments of user position estimation obtained with the same trace of real user movements and noise settings, under three different scenarios of landmark deployment. Specifically, the green curve refers to a scenario in which consecutive landmarks are spaced of 1m along the path, the cyan curve refers to a scenario in which landmark inter-space has been increased to 4m, and finally the black curve refers to a scenario without landmarks. Since the state model assumes that user direction is constant, the transient phases due to the user alignment on the path direction after each direction change are obviousy affected by some fluctuations. Moreover, while the direction estimate works well in all the cases, the accuracy of the position estimates degrade over time and cannot be improved

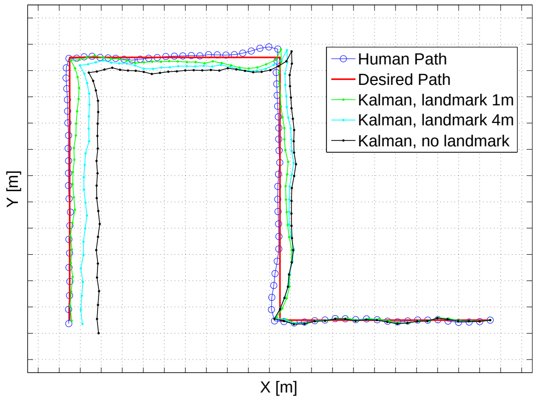

Fig. 7. Kalman-based estimates of human position under different landmark scenarios.

without deploying some landmarks in the environment. Indeed, the user position estimated by the black curve at the end of a path whose overall length is 40m is about 1.5m far from the real position.

V. Conclusions

The ARIANNA system is a solution for helping autonomous navigation of visually impaired people with minimal deployment costs (the colored tapes on the floor) and very simple user interface. Despite of this simplicity, modeling the human walking behavior when navigation is assisted by ARIANNA is a challenging problem. Differently from robot, where feedback decisions can be driven by some state estimates based on optimal filtering of the environmental measurements, user decisions rely on his own perception of the current state and correction strategy.

In the paper, we propose to model human navigation by assuming that user velocity is almost constant and the heading direction is proportionally corrected according to two metrics provided by ARIANNA: the misalignment between the heading direction and the path and the distance from the path. Even when the path is lost, vibration signals guide the user by indicating a circular path oriented in a direction opposite to the one in which the path has been lost, thus allowing to reach the end of the path under various settings of the perception noises and correction coefficients. Since the user always starts navigating from a known point on the path, we also consider tracking the user position during navigation. To this end, landmarks need to be deployed along to path line to avoid the accumulation of position errors.

References

[1] N. Fallah, I. Apostolopoulos, K. Bekris, E. Folmer, The User as a Sensor: Navigating Users with Visual Impairments in Indoor Spaces using Tactile Landmarks, Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems (CHI'12), Pages 425-432, Austin, Texas, May 2012.

[2] S. Willis, S. Helal. RFID information grid and wearable computing solution to the problem of wayfinding for the blind user in a campus environment, IEEE International Symposium on Wearable Computers (ISWC 05), 2005.

[3] N. Fallah, I. Apostolopoulos, K. Bekris, E. Folmer, Indoor Human Navigation Systems; A survey, Interacting with Computers, Oxford Journals. 25:1 Pages 21-33, February 2013.

[4] P. Gallo, I. Tinnirello, L. Giarré, D. Garlisi, D. Croce, and A. Fagiolini, "ARIANNA: pAth Recognition for Indoor Assisted NavigatioN with Augmented perception". Arxiv submit/0867698, 2013.

[5] S.Y. Chen, Kalman Filter for Robot Vision: A Survey, IEEE Transactions on Industrial Electronics, vol.59, no.11, pp.4409-442, Nov. 2012.

[6] S. Munir, J.A. Stankovic, C.J.M. Liang, S. Lin. Cyber Physical System Challenges for Human-in-the-Loop Control. 8th International Workshop on Feedback Computing, USENIX 2013.

[7] S. Munir, J. A. Stankovic, C. J. M. Liang, and S. Lin, "Cyber Physical System Challenges for Human-in-the-Loop Control". In Presented as part of the 8th International Workshop on Feedback Computing. USENIX, 2013.

[8] J. M. Loomis, R. G. Golledge, R. L. Klatzky, Navigation System for the Blind: Auditory Display Modes and Guidance, April, Vol. 7, No. 2, Pages 193-203, MIT press journal, 1998.

[9] Holland, S., Morse, D.R., Gedenryd, H. Audiogps: Spatial audio navigation with a minimal attention interface. Personal Ubiquitous Computing 6, pp. 253-259, 2002.

[10] Etter, R., Specht, M. Melodious walkabout - implicit navigation with contextualized personal audio contents, in: Adjunct Proceedings of the Third International Conference on Pervasive Computing, Munich, Germany, May 8-13, pp. 43-49, 2005.

[11] V. Hayward, O. R. Astley, M. Cruz-Hernandez, D. Grant, G. Robles-De-La-Torre, Haptic interfaces and devices, Sensor Review, ISSN 0260-2288, Feb 2004.

[12] E. Samur, Systematic Evaluation Methodology and Performance Metrics for Haptic Interfaces, PhD Thesis, EPFL Ecole Polytechnique Federale de Lausanne, Switzerland 2010.

[13] Loomis, J. M. Digital map and navigation system for the visually impaired. Unpublished manuscript, Department of Psychology, University of California, Santa Barbara, 1985.

[14] The MoBIC Project, http://isgwww.cs.uni-magdeburg.de/projects/mobic/mobicuk.html

[15] M. Solazzi, W. R. Provancher, A. Frisoli, M. Bergamasco, Design of a SMA Actuated 2-DoF Tactile Device for Displaying Tangential Skin Displacement, World Haptics Conference (WHC), 2011

[16] Amemiya, T., Yamashita, J., Hirota, K., Hirose, M. Virtual leading blocks for the deaf-blind: A real-time wayfinder by verbal-nonverbal hybrid interface and high-density rfid tag space, in: VR '04: Proceedings of the IEEE Virtual Reality IEEE Computer Society, pp. 165, 2004.

[17] G. Park, S. Choi, Perceptual Space of Amplitude-Modulated Vibrotactile Stimuli, World Haptics Conference (WHC), 2011.

[18] W. Heuten, N. Henze, S. Boll, M. Pielot, Tactile Wayfinder: A Non-Visual Support System for Wayfinding, Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges, Lund, Sweden, 2008.

[19] M. Pielot, and R. de Oliveira. Peripheral Vibro-Tactile Displays, ACM Press, 2013. Mobile HCI 2013 - tactile user interfaces. doi:10.1145/2493190.2493197.

[20] M. Pielot, How the Phones Vibration Alarm Can Help to Save Battery Accessed May 12, 2014. http://pielot.org/2012/12/11/how-the-phones-vibration-alarm-can-help-to-save-battery/.

[21] A. Morro, A. Sgorbissa, R. Zaccaria. Path following for unicycle robots with an arbitrary path curvature. IEEE Transactions on Robotics, Vol. 27(5), pp. 1016-1023, 2011.