VIPEye: Architecture and Prototype Implementation of Autonomous Mobility for Visually Impaired People

Waqas Nawaz[1], Khalid Bashir[2], Kifayat Ullah Khan[3], Muhammad Anas[4], Hafiz Obaid[5], Mughees Nadeem[6]

Dept. of Computer and Information Systems, Islamic University of Madinah, Al-Madinah, KSA[1,2]

Fast Data Science (FDS) Lab, Dept. of Computer Science,

National University of Computer and Emerging Sciences (FAST-NUCES), Islamabad, Pakistan[3,4,5,6]

Abstract—Comfortable movement of a visually impaired person in an unknown environment is non-trivial task due to complete or partial short-sightedness, absence of relevant information and unavailability of assistance from a non-impaired person. To fulfill the visual needs of an impaired person towards autonomous navigation, we utilize the concepts of graph mining and computer vision to produce a viable path guidance solution. We present an architectural perspective and a prototype implementation to determine safe & interesting path (SIP) from an arbitrary source to desired destination with intermediate way points, and guide visually impaired person through voice commands on that path. We also identify and highlight various challenging issues, that came up while developing a prototype solution, i.e. VIPEye - An Eye for Visually Impaired People, to this aforementioned problem, in terms of task's difficulty and availability of required resources or information. Moreover, this study provides candidate research directions for researchers, developers, and practitioners in the development of autonomous mobility services for visually impaired people.

Keywords—*Safe and Interesting Path (SIP); Visually Impaired People (VIP); autonomous; mobility; computer vision; path guidance; VIPEye prototype; navigation; graph mining*
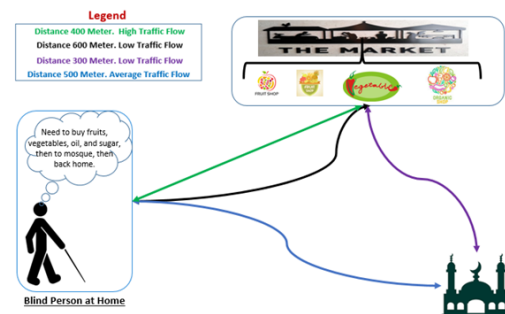
Fig. 1. A blind person wants to go market, then to mosque, and back home. There are two routes to the market. The green colored is short but busy road, whereas, black colored is longer but empty road. Similarly, there are dedicated fruit shops in the market and a combined shop of vegetables and fruits. To provide an optimal solution, the black colored path and a combined shop should be suggested as a navigation path.

I. INTRODUCTION

Outdoor movement of a visually impaired person is limited and hard due to complete or partial short-sightedness [1]. Any non-impaired person can assist them while going out, however, this solution is not always viable. Fortunately, with the recent developments in technology, various innovative methods of assistance have been proposed based on ultrasonic devices, smart cameras, transplantation of robotic eyes, blind navigation system using GSM and RFID and voice navigation system among others [2]. One of the key essence of these developments is to improve the computer-vision technology to fulfill the visual needs of a blind person. To this end, various researchers are using vision based methods [3] focusing issues like sidewalk assessment, zebra crossing spotter, public cataloging of various objects like trees, detecting curb ramps, and public transit accessibility to facilitate smooth and safe walk of blind persons.

Computer vision methods [4]–[8] are also in use for a similar problem of robot vision where the task is to make the robot aware of its surrounding, to help it move around. This is a wide area of research known as Simultaneous Localization and Mapping (SLAM). SLAM uses a range measurement device which may be a laser scanner, sonar, vision or any other capable device. In case of vision based devices for range measurement, the problem is categorized as Visual SLAM (vSLAM) and is an extensively researched topic [5]. Combining vSLAM with topological representations of the environment [7] is an effort to yield improved and robust long-term navigation in real world environments for visually impaired people. However, these solution need improvement in the areas of 3D scene understanding, computations of dense flow scene representations for real time implementations and topological mapping of the environment.

Scene understanding is imperative for smooth navigation based on visual data [9]. This involves range measurement, detection and identification of objects from a scene. Smart phones are equipped with stereo cameras which enables calculating distances to objects [10] for range measurement. Object detection and identification is computationally expensive and traditional approaches fail to produce desired level of accuracy. Recent advances in Computer Vision using Deep Learning methods [11] provide good detection and recognition results but at the cost of computation resources. Another family of detection and recognition frameworks called the Unified (One Stage) frameworks also makes use of deep networks and provide significant improvement in terms of speed. We focus on the application of Unified Frameworks for navigation of the visually impaired. Therefore, these detection and recognition frameworks are not applicable to the problem under consider-

ation.

The integration of scene understanding solutions with proprietary or crowd-sourced geospatial mapping services like Google Maps or OpenStreetMap (OSM)[1] respectively can assists the movement of a blind person from a given source to destination(s) in the outdoor environment. However, we understand that such combination requires further developments in visual technology and investigation over path identification problem. For instance, Google Maps or similar applications suggest the shortest path(s) between source and destination without considering VIP friendly parameters such as higher traffic flow, rush of vehicles, and presence of obstacles. An example scenario is depicted in Fig. 1 for clarity. Finding a suitable path for a pedestrian, having visual dis-ability and considering dynamics of the surroundings, is a challenging task [12]. The multi-criteria shortest path (MCSP) [13] problem and Skyline path computation [14] are similar to determine a suitable path considering various factors. Despite the effectiveness of MCPS and Skyline approaches for general mobility, it is not readily applicable to our problem due to involvement of scene understanding for guidance, which is hard to model.

In this article, we propose a system called SIP-vSUN (Safe and Interesting Path with Visual Scene Understanding) in order to help the visually impaired people smoothly navigate in the surroundings. The prospective solution is to be deployed on a smart phone equipped with the required commodity sensors. The computation heavy load of vision based methods are still not easily handled by smart phones, however, there are preliminary works showing the possibility of using vision based method on smart phones [15]. A real time object detection system called PeleeNet [16] has been proposed for mobile devices. Our proposed system relies on extending the existing works in object detection and recognition for its accuracy and speed and integrating it with range measurement of identified objects thus making it capable of working with required precision on hand held devices and additionally defines the safe and interesting paths for the visually impaired. Our SIP-vSUN system on a smart phone communicates with a blind person through off-the-shelf speech system, which supports voice commands for inputs and voice instructions for guidance and navigation. There are existing libraries like ARKit [17] and ARCore [18] on iOS and Android platforms respectively to realize the implementation of visual components of this system. PyTorch's [19] cross platform libraries will be used for the implementation of the deep network on mobile devices.

The potential contributions of this research work are as follows.

- We are aiming to envision the research community and practitioners on utilize existing tools and techniques, in graph mining and computer vision, towards the development of an effective autonomous mobility system. This system assists visually impaired people, who are deprived naturally or by accident, to walk easily in the surroundings.

- This study gives a clear direction to develop reasonable system, as an application on a smart phone,

which can be used autonomously by visually impaired persons.

- This study explores methods, balanced in their computational needs and performance, in the domain of object detection, recognition, and graph mining.

- The development of a prototype is the realization of our efforts in this domain, which has the potential to guide visually impaired people in a selected scenario and has many expects to be improved in near future.

The rest of the content is organized as follows. Section II explains the proposed methodology where we discuss the notion of safe and interesting path discovery and visual scene understanding for visually impaired people. The design and implementation perspective of the proposed application, i.e. VIPEye, is presented in Section III. Section III-C provided a brief comparative analysis of the the proposed application with the existing similar applications. The potential research directions in this domain are outlined in Section IV and conclusion is drawn in Section V.
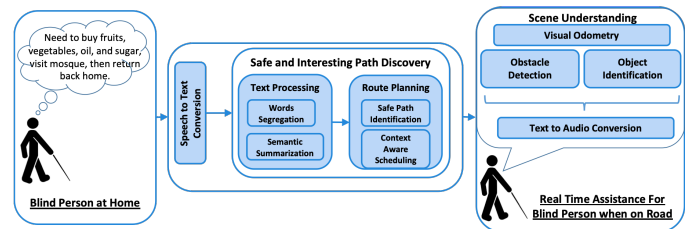

Fig. 2. Proposed System Architecture

II. Safe and Interesting Path with Visual Scene Understanding (SIP-vSUN)

In this section, we present the details of our proposed framework and Fig. 2 depicts the overall system architecture. The processing starts when a person plans to do something like buying items and/or visiting different places. The person talks to the system installed on his smart phone, where off-the-shelf speech to text component of SIP transforms the audio message into text. Next semantic summarization, route planning, and context-aware evaluation using textual information received from the speech engine are performed. The result is a path to be used by the system for navigation. The navigation is assisted through the visual component of our system in terms of real time object detection, recognition and range measurement. When the user is out in the open, the visual and navigation components interact with one another to update the path or guide in real time based on the environmental inputs.

A. Safe and Interesting Path (SIP) Discovery

We now discuss the path planning component of our system. We present its example in Fig. 3 and 4 to deliver the overall idea.

Initially, we transform the obtained audio message into text, as guided in [7]. In order to obtain the action-items, we need to filter the important words from the message. Finally, we perform semantic summarization to group the contextually

---

Algorithm 1: SIP-vSUN Algorithm

---

   input : Destination d, List of Criteria for Path
              Selection $C = \{c_1, c_2, ..., c_n\}$
   output: Comfortable navigation of a Visually
             Impaired Person on SIP

1   Current location selected as starting point s;
    // Candidate Path List
2   Retrieve set of candidate paths $P = \{p_1, p_2, \ldots, p_k\}$
   from s to d using navigation APIs;
    // Identify SIP from candidate paths
3   **for** *each $p_i \in P$* **do**
4   Evaluate p_i for $\forall C$;
        // $weight(p_i)$ is sum of scores for
           path p_i
5   **if** $weight(p_i) ¿ weight(p_j)$ **then**
6   Choose p_i as SIP;
7   **end**
8   **end**
    // Path Navigation
9   $vSUN$ for navigation on SIP;

---

similar nouns together. For instance, fruits, vegetables, oil, sugar, stated in Fig. 3, are eatable items, whereas home and mosque are venues to find path for. In this way, we group the items together into relevant categories. Such grouping then helps to have an optimal path for smooth navigation. Here the challenging issues are Online Analytical Processing (OLAP) style aggregation and vocabulary of locality.
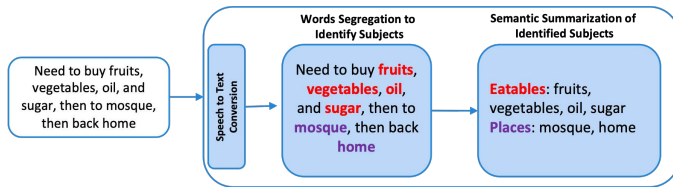

Fig. 3. Text Processing Module

The path identification component comes into action when we have items summarized into relevant grouping. Based on this information, the path discovery module identifies the current location of the person and the places to visit. During this processing, it communicates with 3rd party services (such as Google Maps, OpenStreetMap, MapBox) to determine safe and interesting path among possible set of paths between source and destination. The terms *safe* and *interesting* refer to least number of obstacles and hurdles for VIP mobility and most point of interests covered in short distance by the proposed path. This component also aims to perform clustering of actionable items like identifying the actions which can be performed together. For instance, fruits and vegetables can be purchased together from same shop. Similarly, oil and sugar can be bought together from same grocery store. On the other hand, mosque and home, from the example, are different workable options, so they should not be mixed with shopping agenda. In this way, the system aims to perform hierarchical clustering to identify an interesting path. In our case, we have challenge of hierarchical entity clustering.

The context-aware scheduling is very important in a sense

that a blind person is ignorant of the ground realities happening around. For instance, road construction and maintenance works are in progress, which makes it difficult and less interesting to walk. Similarly, the person is short of time to miss the prayer in the mosque, so he can adjust configuration. In this case, we have multiple criteria to fulfil prior to determine a best route, which is similar to skyline computation in literature. Therefore, the challenge is skyline and multi-criteria path formulation and its efficient computation. For initial experiments and simulations, we expect to utilize the GeoLife - GPS trajectories dataset by Microsoft Research Asia that contains outdoor movements of 17,621 trajectories of duration of 48,000 hours.
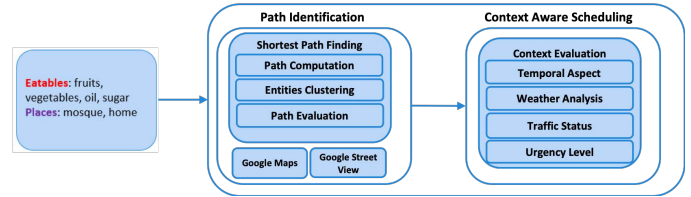

Fig. 4. Route Planning Module

B. Visual Scene Understanding (vSUN)

We present an overall architecture of the Visual Scene Understanding module in Fig. 5. All the inputs required are readily provided by the sensors already available in current smart-phones making it an excellent choice for this application. In addition to the Visual Odometry for range measurement, the proposed system is assisted by incorporating object detection and recognition for better scene understanding. In the following, we present our approach to address aforementioned challenges.
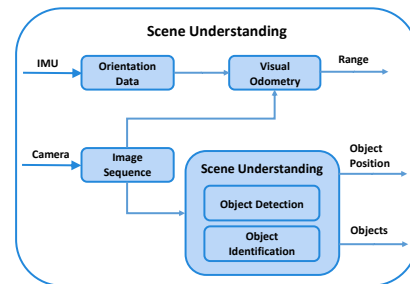

Fig. 5. The block diagram of Scene Understanding Module

Improving scene understanding, with regards to the specific problem of navigation for visually impaired people, involves identification of obstacles, moving objects, pedestrians, sidewalks, zebra crossings and roads. In addition to the identification of static and moving objects, our focus is on tracking moving objects in the scene to handle crowded environments. A better scene understanding [9] eventually benefits localization service for the visually impaired person within the environment.

Detection and identification of object and computation of bounding boxes around them in the scene is a challenging task to be completed in real time. In addition, there are problems associated with real world imagery giving spurious results. To this end, PeleeNet [16] is neural network based solution

for real time object identification and tracking. However, the problem with this approach is that it is trained on standard datasets which have generic classes and do not cover the range of objects including zebra crossing, sidewalks, roads, obstacles etc. In order to address this limitation, we propose to use *transfer learning* in two phases. In the initial phase with limited data, we propose to use the PeleeNet's convolutional network as a fixed feature. This essentially means removing the last fully connected layer and training a classifier on the convolutional codes received from the convolutional network. In the final phase when more data is gathered we propose to initialize the network with pre-trained weights and fine tune the convolutional neural network weights using our own dataset. This phase wise strategy provides the desired level of accuracy initially when there is not much data and improves even further as more data is accumulated. The fine tuning of the network weights can not be done in real time and should be done offline. PeleeNet not only provides us with an advantage of close to real time recognition of objects, it also has a model size which is two thirds of the model size used by similar solutions like MobileNet [20]. Objects identified are used by the Visual Odometery component for range measurement.

III. VIPEye Prototype Design and Implementation

In this section, we elaborate on the prototype [21] as guidance application (a limited version of the actual application) for VIPs. Initially, we discuss the design and implementation details such as the path planning and visual scene understanding towards autonomous mobility along with user interfaces. Afterwards, we provide an abstract comparative analysis on various existing systems and applications developed for VIPs in path guidance context.
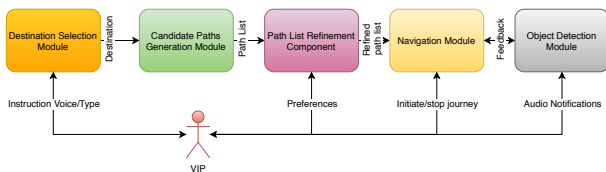

Fig. 6. Main components of our prototype application

A. Design Considerations

There are five main components involve in our implemented prototype application for VIPs. The block diagram shows the sequentially dependent modules in Fig. 6. Initially, the blind person specifies the destination through voice command or typing the name manually with the help of accessibility service in the application, which is the task of destination selection module. We assume that the application is started prior to this step and gets the current location of the user. Once, the destination is selected, the next module determines a set of candidate paths (i.e path list) from source to destination location. We can move from one location to another by following different paths. However, choosing one from the given set of paths is non-trivial and subjective task given the circumstances. Therefore, in the refinement module, we take the preferences from user to filter and choose an appropriate path from path list. Currently, the preferences are limited to various point of interests such as hospital, mosque,

school, university, etc. Multi-criteria shortest path finding is a well studied problem and is used as a candidate solution for path selection based on user preferences. Navigation and object detection modules can work and coordinate together once the user starts navigation on the selected path. These modules notify user in terms of standard audio messages similar to general purpose navigation applications. The object detection module covers limited set of objects in this prototype and subject to cover diverse kind of objects in next version.

The choice of an appropriate interaction interfaces for visually impaired users is also critical. Our prototype supports accessibility option with native support of the Android platform and designed simple interfaces to increase its usability for VIPs. User interfaces of our prototype with visual and non-visual aspects are presented in Fig. 8 and 7, respectively. Notice that our prototype application also allows the user to manually capture an image and run a deeper model on a better quality image for an improved scene understanding. The default operation of the application is to automatically capture images every $1/4_{th}$ of a second and run it through a standard model to meet time constraints.

B. Development Perspective: Libraries and Technologies

We dedicate this section to elaborate on development aspects of our prototype system [21] as a mobile application implemented on Android platform.

The collection of static model elements such as classes, types, contents, and their relationships is presented in Fig. 9. The important functions of the project include *OnActivityResult* (to save the entered string as a destination), *getRoute* (to get the route that leads from source to destination), *getRoutecoordinates* (to save the whole route that leads from source to destination), *makeGeocodeSearch* (searches for places on route that leads from source to destination), *updateRoute* (updates the multiple routes leading from source to destination to a best route that leads from source to destination), *takePicture* (to take picture of objects that comes in user's path), *detectObject* (to check that the picture taken belongs to which category), and *speak* (To speak whatever it is passed to the function as an argument such as places names).

We briefly highlight important activities involved in our application that include path finding, navigation and object detection. In order to find path on user's way, the user enters the destination, the application checks whether the destination is valid or not. If valid then it asks to enter criteria and calculates best path for navigation. Else it asks the user again to enter the destination. To navigate on the path that leads to destination from source, the user selects the path that the application has returned in the activity called path finding. Our application tells user that path has been selected and also inform the estimated time to reach the destination. The user starts the journey and application guides the user in that journey on turn basis to reach the destination. The object detection activity involves the scene capturing and understanding. To detect objects, that comes in user's path, our application takes the picture of the scene with objects for every 1/4th of a second, categorize that picture and notifies the user if their is a detected obstacle on the way. User has an option of getting a more information about the obstacle by pressing the capture button as shown in Fig.
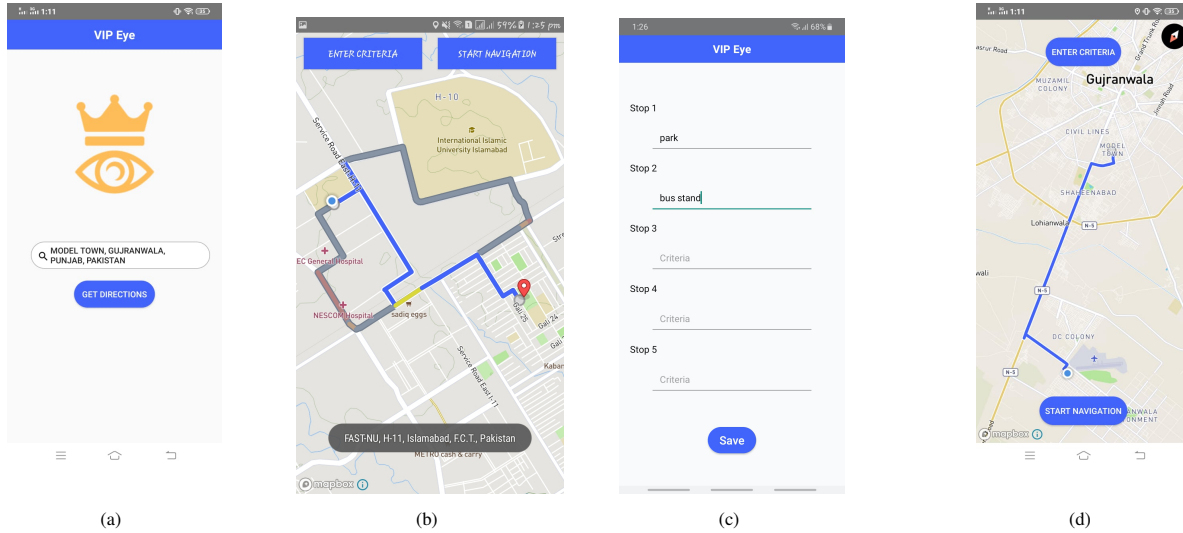
(a)   (b)   (c)   (d)

Fig. 7. Prototype User Interfaces (Non-Visual Aspect) (a) Home Screen (b) Multiple routes for a selected pair of location (c) Preferences for mobility (d) Selected Path

(a)   (b)   (c)

Fig. 8. VIPEye Prototype Navigation and Obstacle Detection (Visual Aspect) with Audio-Feedback (a) Navigation (b) Manual Fixed Obstacle Detection (Detailed information) (c) Manual Movable Stationary Obstacle Detection (Detailed information)

8 allowing a more complex model to run on a better quality image and extract detailed information about the objects in the scene. After the completion of this operation user can then continue with the assisted navigation.

- **Main Page:** The application starts with a search bar in which the user enters the destination. This search bar has autocomplete feature that shows names of places that match with the name that the user has entered. We used MapBox platform for this purpose and considered various libraries such as Mapbox GeocodingCriteria, Mapbox CarmenFeature, and Mapbox PlaceAutocomplete.
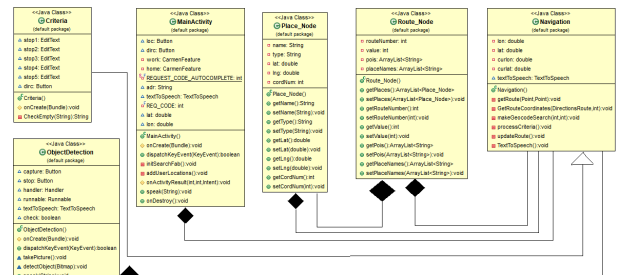

Fig. 9. Class diagram of the prototype implementation

```
1  public void onClick(View view)
2  {
3    Intent intent = new PlaceAutocomplete.
   IntentBuilder().build();
4    startActivityForResult(intent,
   REQUEST_CODE_AUTOCOMPLETE);
5  }
6
```

Listing 1: OnClickListner pseudo code for the search box

- **Navigation Page:** The navigation page shows multiple paths leading to the destination (without any criteria). From here the user goes to the criteria page and enters up-to five criteria places in order of priority. Then the navigation page shows the best chosen path based on criteria.

```
1  private void getRoute(Point origin, Point
    destination)
2  {
3    NavigationRoute.builder(this).origin(
   origin).dest(dest).build().getRoute();
4  }
5
```

Listing 2: Getting routes from origin to destination

Then using the above function we also get the places on each route using makeGeocodeSearch function. By pressing the Start Navigation button the application guides the user to their destination. And by pressing the Start Walking button the user is be able to detect obstacles in their path. We achieved this functionality through MapBox libraries such as Mapbox LatLng, Mapbox MapView, Mapbox MapboxMap, Mapbox Style, Mapbox NavigationLauncher, Mapbox NavigationRoute, and Mapbox DirectionsRoute.

- **Navigation and Obstacle Detection Page:** The obstacles detection page shows the live stream, the current photo and the obstacles detected in the current photo. We used OtaliaStudios CameraView and Firebase FirebaseVisionObjectDetector for navigation and object detection purposes.

```
1  /*Live Stream*/
2  cameraView = findViewById(R.id.
   id_cameraview);
3  cameraView.setLifecycleOwner(this);
4  //Take Picture (when user presses the
   volume up key)
5  cameraView.takePicture();
6  cameraView.addCameraListener(
   onPictureTaken(result){detectObjects(
   result); });
7
8  /*Object Detection*/
9  void detectObjects(result)
10 {
11   objectDetector.processImage(image) {
12     public void onSuccess(List<
   FirebaseVisionObject> detectedObjects){
13     // Displays the name(s) of the detected
    object(s)
14     }};
15 }
16
```

Listing 3: Partial code on navigation and object detection

- **Non-visual Aspects:** It is obvious that VIPs are unable to see the interfaces of our proposed application rather communicate with it via non-visual senses such as speaking and hearing. Therefore, we provided an audio-based interaction and feedback mechanism implemented through existing libraries, i.e. AndroidSpeech TextToSpeech and AndroidSpeech's SpeechRecognizer, to covert from audio-to-text and vice verse.

```
1  /*Text-To-Speech*/
2  public void speak(String data)
3  {
4    textToSpeech.speak(data,TextToSpeech.
   QUEUE_FLUSH,null);
5    // can be modified as per requirements
6  }
7  /*Speech-To-Text*/
8  Intent intent = new Intent(
   RecognizerIntent.ACTION_RECOGNIZE_SPEECH)
   ;
9  // more options can be added to this
   function
10 startActivityForResult(intent, REQ_CODE);
11 void onActivityResult(int requestCode,
   int resultCode, Intent data) {
12   // Check requestCode and resultCode for
    confirmation
13   ArrayList result =
14   data.getStringArrayListExtra(
   RecognizerIntent.EXTRA_RESULTS);
15   // Display or perform action according
   to what was spoken
16 }
17
```

Listing 4: Non-visual interaction mechanism for VIPs

C. Comparative Analysis

We describe few mostly related applications and systems used to assist in the navigation of the blind people as also highlighted in [22]. Recent surveys [23] [24] [25] [26] discussed various kinds of systems and applications developed for visually impaired people. These systems and applications are either the outcomes of research work in this domain as prototypes or entirely as business oriented products. These solutions are categorized into general purpose object recognition based [27]–[32], navigation related [33]–[41], and specialized systems and devices [42]–[46]. The discussion in this section highlights key aspects these systems cover in terms of guiding visually impaired people and also provide insights to improve them for better guidance.

We provide an overview of comparative analysis for various applications and systems with respect to predefined criteria in Table I. The criterion is selected based on suitability of these applications and systems for visually impaired people towards autonomous navigation, which is the main focus of this article. GPS criterion depicts whether the application or system supports location-aware services or not. These applications interact with the users through either audible instructions or haptic feedback. Not all applications and systems analyze images and video data to observe the surroundings and develop an understanding for the VIPs. The understanding of the surroundings depends upon effectively detecting and identifying the objects. Object detection is a key step towards scene understanding, where the majority of the techniques under

consideration detect various kinds of objects from individual images. However, few approaches detect objects continuously from a series of images taken through built-in camera. Most of the applications are available free of cost on either one or both platforms, i.e. iOS or Android. It is critical for any application or system to be autonomous so that a visually impaired person feels more independent. This is one of the major factors to influence the usability of any system or application, as depicted in the last column of the comparison table. It is interesting to know that existing solutions are either more focused on non-visual information, e.g. GPS signals and navigational semantics, or visual content to understand the surroundings. In our understanding, visual content analysis is a complex task compared with the analysis of non-visual information towards navigational guidance, therefore, many aspects in visual content analysis are still under consideration.

IV. Research Directions

The potential research directions, while developing an effective solution for visually impaired people towards autonomous navigation, are as follows.

1) SIP prediction algorithm and repository: To develop an effective algorithm to predict or identify the safe and interesting paths, which are eventually stored in a repository for future correspondence. The identification of the SIPs through experience will result in a rich repository.

2) Intelligent path discovery approach: An approach needs to be developed to discover or explore paths from given source to destination(s). The path discovery problem emphasizes on various criterion defined or customized for impaired person. Defining the representative criterion for impaired person is also major concern.

3) Hazard detection and identification in real-time: We need an algorithm to detect various kinds of obstacle/hazards instantly to take respective measures by the impaired person while navigating known/unknown territory. This algorithm expected to utilize vital information (visual & non-visual) to make decisions quickly. The identification of such a useful information is itself a challenging task.

4) Context-aware navigation system: A novel approach is required to determine the context from the environment through visual and non-visual data. It helps the users to navigate with contextual information (knowledge of surroundings like passing by superstore, ATM, gas-station, etc.).

5) A comprehensive system with smart navigation: This product will be an integration of other essential components together to solve the navigation problem for visually impaired people. This system provides a comprehensive solution to smartly/intelligently navigate through known/unknown territories. This solution will be realized on smart devices (e.g. Smartphones) as these devices are already equipped with essential tools and commodity sensors.

V. Concluding Remarks

In this work, we have conceptualized an overall system as a candidate solution for mobility of visually impaired people in terms of visual scene understanding and graph mining approaches. A prototype implementation is presented and improvements are in progress on algorithms determining safe and interesting path by considering multiple factors associated with blind people. Additionally, We have highlighted various research directions for developers and practitioners to build effective services towards mobility of visually impaired people.

Acknowledgment

References

[1] R. R. Bourne, S. R. Flaxman, T. Braithwaite, M. V. Cicinelli, A. Das, J. B. Jonas, J. Keeffe, J. H. Kempen, J. Leasher, H. Limburg *et al.*, "Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis," *The Lancet Global Health*, vol. 5, no. 9, pp. e888–e897, 2017.

[2] M. M. Islam, M. S. Sadi, K. Z. Zamli, and M. M. Ahmed, "Developing walking assistants for visually impaired people: A review," *IEEE Sensors Journal*, vol. 19, no. 8, pp. 2814–2828, 2019.

[3] R. Jafri, S. A. Ali, H. R. Arabnia, and S. Fatima, "Computer vision-based object recognition for the visually impaired in an indoors environment: a survey," *The Visual Computer*, vol. 30, no. 11, pp. 1197–1222, 2014.

[4] P. F. Alcantarilla, "Vision based localization: from humanoid robots to visually impaired people," *Electronics (University of Alcala, 2011)*, 2011.

[5] T. Taketomi, H. Uchiyama, and S. Ikeda, "Visual slam algorithms: a survey from 2010 to 2016," *IPSJ Transactions on Computer Vision and Applications*, vol. 9, no. 1, p. 16, 2017.

[6] R. Tapu, B. Mocanu, and T. Zaharia, "Wearable assistive devices for visually impaired: A state of the art survey," *Pattern Recognition Letters*, 2018.

[7] L. F. Rojo, L. Paya, F. Amoros, and O. Reinoso, "Using global appearance descriptors to solve topological visual slam," in *Advanced Methodologies and Technologies in Artificial Intelligence, Computer Simulation, and Human-Computer Interaction.* IGI Global, 2019, pp. 1127–1140.

[8] E. Arnold, O. Y. Al-Jarrah, M. Dianati, S. Fallah, D. Oxtoby, and A. Mouzakitis, "A survey on 3d object detection methods for autonomous driving applications," *IEEE Transactions on Intelligent Transportation Systems*, 2019.

[9] B. Yang, Z. Lai, X. Lu, S. Lin, H. Wen, A. Markham, and N. Trigoni, "Learning 3d scene semantics and structure from a single depth image," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 309–312.

[10] Y. D. Salman, K. R. Ku-Mahamud, and E. Kamioka, "Distance measurement for self-driving cars using stereo camera," in *proceeding 6th Int. Conf. Comput. informations.* Universiti Utara Malaysia, 2017, pp. 235–242.

[11] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen, "Deep learning for generic object detection: A survey," *arXiv preprint arXiv:1809.02165*, 2018.

TABLE I. COMPARISON OF VARIOUS PATH GUIDANCE APPLICATIONS AND SYSTEMS IN TERMS OF THEIR CAPABILITIES AND THE QUALITY OF SUPPORTED FUNCTIONALITY.

Apps & Systems	GPS	Audio or Haptic	Image	Video	Object Detection	Free	Autonomous	Usability
NavCog3 [33]	No	Yes	No	No	No	Yes	Yes	MED
WayFindr [35]	No	Yes	No	No	No	Yes	Yes	MED
Be My Eyes [27]	No	Yes	Yes	Yes	No	Yes	No	MED
TapTapSee [28]	No	Yes	Yes	No	Yes	Yes	Yes	LOW
iDentifi [30]	No	Yes	Yes	No	Yes	Yes	Yes	LOW
SeeingAI [31]	No	Yes	Yes	Yes	Yes	Yes	Yes	MED
BeSpecular [32]	No	Yes	Yes	No	No	Yes	No	LOW
Microsoft soundscape [34]	Yes	Yes	No	No	No	Yes	Yes	LOW
BlindSquare [37]	Yes	Yes	No	No	No	No	Yes	MED
Lazarillo [39]	Yes	Yes	No	No	No	Yes	Yes	LOW
Intersection Explorer [36]	Yes	Yes	No	No	No	Yes	Yes	MED
Lazzus [40]	Yes	No	No	No	No	Yes	No	LOW
Arianna Navigatio [38]	No	Yes	No	Yes	Yes	Yes	No	LOW
NaviLens [41]	No	Yes	Yes	No	Yes	Yes	Yes	MED
VIP-Eye (Prototype) [21]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	HIGH
Sunu [42]	No	Yes	No	No	Yes	No	Yes	LOW
OrCam [45]	No	Yes	Yes	No	Yes	No	No	LOW
Blaid [46]	Yes	Yes	Yes	No	Yes	No	Yes	MED
eSight [43]	No	Yes	Yes	No	Yes	No	Yes	MED
iBionics [44]	No	Yes	Yes	No	Yes	No	Yes	MED

[12] J. E. Froehlich, A. Brock, A. Caspi, J. Guerreiro, K. Hara, R. Kirkham, J. Schöning, and B. Tannert, "Grand challenges in accessible maps," *Interactions, Association for Computing Machinery*, 2019.

[13] N. Shi, S. Zhou, F. Wang, Y. Tao, and L. Liu, "The multi-criteria constrained shortest path problem," *Transportation Research Part E: Logistics and Transportation Review*, vol. 101, pp. 13–29, 2017.

[14] K. Deng, X. Zhou, and H. Tao, "Multi-source skyline query processing in road networks," in *2007 IEEE 23rd international conference on data engineering*. IEEE, 2007, pp. 796–805.

[15] T. Y.-H. Chen, L. Ravindranath, S. Deng, P. Bahl, and H. Balakrishnan, "Glimpse: Continuous, real-time object recognition on mobile devices," in *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*. ACM, 2015, pp. 155–168.

[16] R. J. Wang, X. Li, and C. X. Ling, "Pelee: A real-time object detection system on mobile devices," in *Advances in Neural Information Processing Systems*, 2018, pp. 1963–1972.

[17] "Arkit 3: Get ready for the latest advances in augmented reality," 2020. [Online]. Available: {https://developer.apple.com/augmented-reality/}

[18] "Arcore: With arcore, build new augmented reality experiences that seamlessly blend the digital and physical worlds." [Online]. Available: {https://developers.google.com/ar/}

[19] "Pytorch: An open source machine learning framework that accelerates the path from research prototyping to production deployment." [Online]. Available: {https://pytorch.org/}

[20] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.

[21] "Visually impaired persons' eye (vip-eye): Safe and smooth navigation for visually impaired people using smart phone." [Online]. Available: {https://github.com/ASHIR14/VIP-Eye}

[22] NoisyVision.org, "Apps and devices for blind and visually impaired people." [Online]. Available: {https://www.noisyvision.org/2018/08/01/app-and-devices-for-blind-and-visually-impaired/}

[23] R. Tapu, B. Mocanu, and T. Zaharia, "Wearable assistive devices for visually impaired: A state of the art survey," *Pattern Recognition Letters*, 2018.

[24] W. Elmannai and K. Elleithy, "Sensor-based assistive devices for visually-impaired people: Current status, challenges, and future directions," *Sensors*, vol. 17, no. 3, p. 565, 2017.

[25] S. Real and A. Araujo, "Navigation systems for the blind and visually impaired: Past work, challenges, and open problems," *Sensors*, vol. 19, no. 15, p. 3404, 2019.

[26] A. Darvishy, "Assistive technologies: short overview and trends," in *1st Swiss conference on barrier-free communication, Winterthur, 15-16 September 2017*. ZHAW Zürcher Hochschule für Angewandte Wissenschaften, 2018, pp. 52–55.

[27] H. J. Wiberg, "Be-my-eyes: Bringing sight to blind and low-vision people." [Online]. Available: {https://www.bemyeyes.com/}

[28] "Taptapsee: Assistive technology for the blind and visually impaired." [Online]. Available: {https://taptapseeapp.com/}

[29] "Cloudsight: Image processing library." [Online]. Available: {https://cloudsight.ai/}

[30] "identifi: Mobile application for partially sighted and visually impaired people." [Online]. Available: {http://getidentifi.com/}

[31] A. Seeing, "Talking camera app for those with a visual impairment," *Retrieved September*, vol. 15, 2018. [Online]. Available: {https://www.microsoft.com/en-us/ai/seeing-ai}

[32] "Bespecular: Let blind people see through your eyes." [Online]. Available: {http://www.bespecular.com/en/}

[33] D. Sato, U. Oh, K. Naito, H. Takagi, K. Kitani, and C. Asakawa, "Navcog3: An evaluation of a smartphone-based blind indoor navigation assistant with semantic features in a large-scale environment," in *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, 2017, pp. 270–279.

[34] "Microsoft soundscape: A map delivered in 3d sound." [Online]. Available: {https://www.microsoft.com/en-us/research/product/soundscape/}

[35] "Wayfindr: Indoor audio wayfinding and audio navigation." [Online]. Available: http://www.wayfindr.net/wp-content/uploads/2018/08/Understanding-Audio-Wayfinding-and-Audio-Based-Navigation.pdf

[36] "Intersection explorer: Intersection explorer helps blind users explore neighborhoods." [Online]. Available: {https://www.appbrain.com/app/intersection-explorer/com.google.android.marvin.intersectionexplorer}

[37] "Blindsquare: Pioneering accessible navigation – indoors and outdoors." [Online]. Available: {https://www.blindsquare.com/}

[38] E. T. from the arXiv, "App turns smartphone into virtual cane for the blind." [Online]. Available: {https://www.technologyreview.com/s/523401/app-turns-smartphone-into-virtual-cane-for-the-blind/}

[39] "Lazarillo: Orientation and autonomy for people with visual disability." [Online]. Available: {https://lazarillo.cl/en/}

[40] "Lazzus: Navigate the city and explore your surroundings independently." [Online]. Available: {http://www.lazzus.com/}

[41] "Navilens: A brand new system of bidimensional, or bidi, codes." [Online]. Available: {https://www.navilens.com/}

[42] "Sunuband: It's your world. explore it with the sunu band." [Online]. Available: {https://www.sunu.com/en/index.html}

[43] "esight: We make glasses that let the visually impaired see." [Online]. Available: {https://www.esighteyewear.eu/}

[44] "ibionics diamond eye: How a local startup is building a better bionic eye." [Online]. Available: {https://ottawacitizen.com/news/local-news/a-clear-view-how-a-local-startup-is-building-a-better-bionic-eye}

[45] "Orcam myeye 2: The most advanced wearable assistive technology for the blind and visually impaired, provides independence by conveying visual information, audibly." [Online]. Available: {https://www.orcam.com/en/myeye2/}

[46] "Toyota project blaid: Wearable technology to help the blind and visually impaired." [Online]. Available: {https://www.toyota.co.uk/world-of-toyota/stories-news-events/toyota-project-blaid}