

Illumination – The Navigation System Associated with Speech and Image Recognition Based on Location Service

Zhenyu Liu, Rong Bi, Jiawu Li, Heng Shen, Ye Song, Fang Zhao

School of Software Engineering, Beijing University of Posts and Telecommunications, Beijing, China
lzysunny0103@gmail.com

Abstract—this paper deals with the current navigation system for the visual-impaired people, developing an indoor and outdoor navigation system, Illumination, in association with speech and Image Recognition for them. For human-machine interaction, Illumination is launched by shaking the phone instead of traditionally touch the specific icon. All operations for page switch are strictly normalized into four gestures heading for up, down, left and right directions. Furthermore, it is the poor vision of our users that disables them to walk out alone because of the obstacles. By simulating the partial functions of the users' eyes, Illumination utilizes camera to implement obstacle recognition with real-time image. What is more, Illumination designs and implements a new kind of map, list map. All the information about maps is saved and displayed as tree data structure, catering best to the needs of our target users. (*Abstract*)

Keywords- navigation; speech; image recognition; human-machine interaction; list maps (*key words*)

I. INTRODUCTION

The robust system of the prevalent navigation engine such as Baidu Maps and Google Maps has almost satisfied everybody except for the visual-impaired by virtue of the deficiency of their poor eyesight. The improvement of the living standards also stimulates the desire of the visual-impaired people to travel, which makes it vital important for the manufacturer to increase the precision and practicability of the navigation system. Therefore, this paper defines a multi-function navigation system, the one incorporating the speech synthesis and recognition and image recognition. The simple unified gesture and fluent speech synthesis and recognition act as the main form of human-machine interaction. Besides, the image recognition will be used as subsidiary module when running the navigation. We believe our research could make contribution to the research on navigation.

This paper is divided into five parts. The first part is the introduction of Illumination; the second demonstrates the requirement of target users collected in the research; the third elucidates the specific human-machine interaction under the influence of the particular requirement; the fourth illustrates the theory of indoor positioning and the design and implementation of list map; the fifth analyze the data collected from test to assess the performance; the sixth is the conclusion of this paper.

II. RELATED WORK

Having communicated with Chinese Braille Press, we obtain many efficacious characteristics of visual-impaired people. First of all, the deflection in vision of them does develop a fabulous sensitivity of hearing. According to the statistics, the speed of their hearing is three to five times faster than the one of ordinary hearing.[1] Then, on the basis of test data, these people are inclined to use touch as trigger events so that list map, to some extent for these disable, is the best substitute for ordinary map. List map, a map recorded on the tree structure, classifies the information with the hierarchy of the point on the map and saves it in the relative tree nodes. The father nodes must include the information in the child node. Subsequently, the modification on the navigation system is targeted on the hardware and current products intentionally head for high technology and expensive cost. Thus, Chinese Baillie Press wants to design the product born out of the conviction of software to associate the previous products to get more accurate results.

According to the statistics, over 60 percent people out of 6 billion in the whole world have the problem of chronicle eye diseases or visual impairment, in which 9 percent are serious patients or totally blind. [3]China has the most blind in the world – about 5 million, accounting for 18 percent blind in the world. Additionally, there are also over 6 million people with low vision and about 10 million children with strabismus and amblyopic. These potential users group mentioned above almost rely on the public organizations such as Chinese Baillie Press. Low reliability and flexibility of traditional tactile sticks fail to fulfill the requirements and cope with the complicated traffic conditions.

Taking all the factors mentioned above into consideration, we have Illumination tailor made for visual-impaired people. Technologically speaking, it associates image recognition with speech recognition and synthesizer; meanwhile, it supports a new style of map, list map.

III. HUMAN-MACHINE INTERACTION

Illumination emphasizes human-machine interaction to facilitate target users.

A. Voice Interaction

Speech is one of the methods that interact between system and users. The source package of Corporation Xunfei, the

one with the best speech engine in Chinese pronunciation, is applied to implementing the function.

Illumination has encapsulated two main functions, speech recognition and synthesization. Speech recognition, regarded as the main way to input for uses, initializes a series of results saved in a string array. This array will sort in an ascending order by the reliability of recognition results and we only automatically pick up the result with highest reliability to return a string in Chinese. The next step is to transform this string into a string coded by pronunciations,[4]

- Transform the raw string into a hashset, where all the pronunciations inside will take one position, including the polyphones
- Do permutation and combination to redevelop the string in Chinese based on the hashset, and use the semantics to identify all the correct string

After we artificially found our own base for characters, this matching process will be accelerated.[5][6]

B. Gesture Design

User Interface is one of the main characteristics of the smartphone, and it is this one that visual-impaired people feel awkward to operate due to user-unfriendly interface design and complicated operation. On account of contemporary hardware configured on the phone, the gesture adapts to the users' requirement.

Born out of the conviction of concision, we pick out five gestures – Up, Down, Left, Right, and DoubleClick – to composite the elementary gesture operations. Further, through the traverse by level in tree data structure, we are able to define the specific meaning of every move.

- Left is defined as the previous step, and Right is defined as the next step.
- Up and Down are targeted on the shift among all the elements at the same depth in the tree
- DoubleClick is sanctioned as the confirmation of results.

This upgrade is supplied to approximately all the modules of Illumination, actually escalated the User Experience.

IV. OBSTACLE RECOGNITION

Illumination, to some extent, intends to substitute the camera for eyes by the technology of Image recognition. In the system, we come up with a method named “No-training Obstacle Recognition”, which implements recognition without existing data support. Firstly, the system open the camera in order to deal with the vary percentage of two pictures whose duration of capturing are only two frames. Secondly, a threshold value will be configured to compare it with the percentage calculated in the first step; once the percentage exceeds the threshold, the system considers that there is obstacle in front of the user. Having detected the obstacles, Illumination will automatically inform the user with speech to bypass it. The following is the actual algorithm for detection.

- Transform the raw picture into grey scale picture.

- Do Gaussian Smoothing for once to eliminate the noise
- Extract the contour of picture with Canny operator
- Extract the vertexes of the polygons to compare them with the previous picture after recognizing polygons. Equivalently, the system sets up a threshold to judge if there is an obstacle.

*Plus, the relative formula – percentage = current vertexes / previous vertexes

After we do the test on this module, the system get the result that reflects at the beginning and ending.

Beginning,

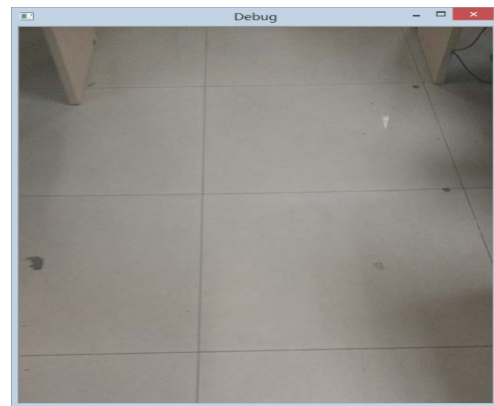


Figure 1. Beginning

TABLE I. THE PARTIAL PARAMETERS AT THE BEGINNING,

Variable	Definition	Value	Ratio
Threshold	The threshold value defined in Image recognition	120	-
Dots	The ratio of current vertexes to previous vertexes	0/0	0.0%
LastDots	Final dots	0	-
differ	The differences between these two computation	0	-
Perimeter	The ratio of the area	0.0/0.0	0.0%

The next,



Figure 2. The ending

TABLE II. THE PARAMETERS AT THE NEXT,

Variable	Definition	Value	Ratio
Threshold	The threshold value defined in Image recognition	120	-
Dots	The ratio of current vertexes to previous vertexes	372/525	70.86%
LastDots	Final dots	0	-
differ	The differences between these two computation	372	-
Perimeter	The ratio of the area	1830.76/3952.79	46.32%

V. INDOOR POSITION AND LIST MAP

A. Inertial Positioning

The inertial sensor and WI-FI signal are applied to positioning on EYE. The inertial positioning presumes the current position with start location, velocity and direction of users. As the following picture shows, the user is located on the start position $L(t_0, x_0, y_0)$ at time t_0 , and move to new position $L_1(t_1, x_1, y_1)$ along with the direction α . Consequently the position L_1 at time t_1 will be:

$$L(t_1, x_1) = L(t_0, x_0) + (t_1 - t_0) \times v \times \sin \alpha \quad (4)$$

$$L(t_1, y_1) = L(t_0, y_0) + (t_1 - t_0) \times v \times \cos \alpha \quad (5)$$

In the formula above, v symbolizes the velocity of user and α represent the angle between the direction of movement and the north. Similarly, we also get the position at time t_2 in the following picture. [5]

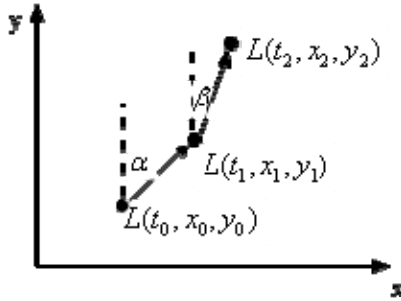


Figure 3. Prediction Sample with Inertial method

Absolutely, by virtue of the cumulative error, the accuracy will go down once the time becomes longer, but its successive real-time positioning is the advantage. The indoor position based on WI-FI signal, though the precision depends on the number and density of trained samples, avoid the cumulative error in a long time. Therefore, the system combines these two approaches to design our function.

B. List Map

The main problem of list map is the transformation from standard map to list map. Then we do this transformation in the following three steps.

Step1: In the partial maps, we regard the present position in general maps as our root node in the data structure. Having finishing the switch, we select all the points marked

on the maps as the child node on the level one with no priority.

Step2: Later on, we should filter which nodes on level one containing indoor space to decide if we dig up the subordinate. In the subsequent aggregation, we pick up the floor as the classification standard to differentiate the block of POIs. At each node on the level one, we specify these nodes in the building as the child node on the level two.

Step3: At the specific floor, we simply list all the POI marked on the maps as the relative child node on the level three.

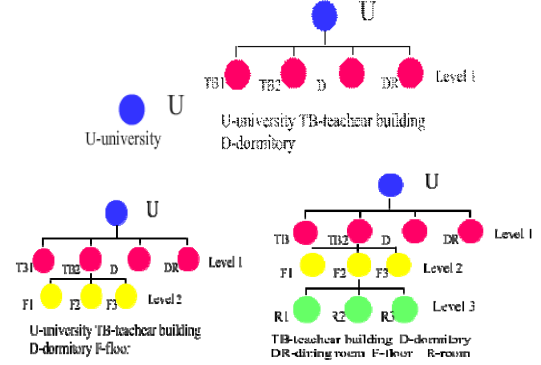


Figure 4. The steps of list map

C. The Algorithm

Through the idea illustrated above, the algorithm of implementing list map will be divided into several steps as followings:

- Initialize the data structure, such as the partial block, vertex, floor, room and so on
- Sort all the nodes according to each level with ascending order. Plus, the nodes with same level will be counted from left to right and put into a node array.
- Create an auxiliary integer array to record the number of child nodes of the father node with the same index in the node array. If the node is the leaf node, the value in the integer array is assigned as 0.

VI. PERFORMANCE ANALYSIS

According to the testing and statistics of positioning technology, approximately 15050 uAh has been consumed in ten minutes, accounting for 7 percentages of the volume of the whole battery.

As for the frequency and accuracy of positioning, in the outdoor navigation, the system positions per second or every meter that the user move from the initial point. No matter which condition has been triggered, the system will start positioning. Besides, the accuracy of positioning is ranged from ten meters to twenty five meters. As for the indoor navigation, the system obtains position information once in every three seconds; and the accuracy is less than five meters.

VII. CONCLUSION

This paper deals with the design and implement of the system, Illumination, associated with speech recognition and synthesization and image recognition. Additionally, based on our system, we could improve the availability of navigation and optimize user experience.

REFERENCES

- [1] Beijing Union Medical University Press, "Health Statistics Yearbook 2012", China
- [2] http://www.who.int/dg/speeches/2012/blindness_20120625/zh/
- [3] The Editorial Department of Health Statistics Yearbook, "Health Statistics Yearbook 2011", China
- [4] <http://blog.csdn.net/pathuang68/article/details/6692882>
- [5] <http://open.voicecloud.cn/index.php/default/doccenter/doccenterInner?itemTitle=eXlqcw==&anchor=Y29udG10bGUyNg==>
- [6] Li Baoxiang, The research on the problems of indexing the keyword of speech, Beijing University of Posts and Telecommunications, Beijing, China, 2013.4.5
- [7] Tiankun Hu, Design and Realize an Indoor Location System Based on Android, Donghua University, China