Proposing a Hybrid Tag-Camera-Based Identification and Navigation Aid for the Visually Impaired

Yasser Ebrahim, Wegdan Abdelsalam
Dept. of Computing & Information Science
University of Guelph,
Guelph, ON, Canada
(ybrahim, wabdelsalam)@uoguelph.ca

Abstract—In this paper a barcode-based system to help the visually impaired identify objects in the environment and navigate through unknown territories is introduced. The system is based on the idea of tagging the different objects with 2D barcodes. With the aid of a portable camera and a computing device, the system can recognize and relay the barcode content to the user. We present the advantages of the proposed system compared to those of existing technologies. The different steps for recognizing and extracting the barcodes are described and applied to a sample image.

Keywords—barcode; visually impaired; navigation; object identification.

I. INTRODUCTION

When the visually impaired interact with their environment, they face two principal challenges. The first is to sense and identify the objects surrounding them. We call this the identification problem. Secondly, the current location as it relates to the destination and route need to be identified [1]. We call this the navigation problem. Aside from the many attempts to allow the visually impaired person to actually see by connecting an artificial retina to the brain, there have been many other approaches that are less invasive. The idea is for the system to identify the user’s surroundings and communicate this knowledge to him/her via some medium.

Regardless of the medium that is adopted to relay the information to the user, there are three groups of systems for helping the visually impaired person with the identification and navigation problems. The three groups differ in the approach that is used to collect the information from the environment. The first approach involves the use of a GPS device to identify the location of the user. By feeding this location to a GIS system, the components of the environment other navigational information can be identified. The second approach is the tagging of the objects in the environment by deploying a set of transmitters whose signals are received by a mobile device the visually impaired person carries. Then, the signal can be used to identify the object that emitted it. The third approach is based on the use of one or more portable camera(s) by which images of the environment are captured. After those images are analyzed, the useful information is extracted and relayed to the user.

In this paper we present a hybrid tag-camera-based system that capitalizes on the advantages of both approaches so that the drawbacks of each approach are minimized.

II. THE STATE OF THE ART

In this section, we introduce a set of systems that represent the three principal approaches to solve the identification and navigation problems for the visually impaired.

A. GPS-based Systems

Since the mid 1980s, GPS-based systems to help the visually impaired have been the focus of researchers [1]. The concept is simple: take the location of the user as reported by the GPS device, query a GIS system by using this location, and relay the information provided by the GIS system to the user. This information can then be integrated for identification (e.g., what building the user is standing next to) or navigation (e.g., how far a certain destination is and in which direction).

Some early GPS-based systems include Strider from Arkenstone of Sunnyvale, California [2], the Electronic Guide Dog project undertaken by a European consortium, the MoBIC project carried out by a UK/Swedish/German consortium [3], and the system developed in Japan by Makino and his colleagues [4]. One of the earliest attempts to implement such a system is the Personal Guidance System, developed by Loomis and his colleagues at the University of California, Santa Barbara [1].

One of the greatest weaknesses of GPS-based navigation systems is the fact that they can only be used outdoors. Indoors or in areas with tall buildings (creating the canyon effect), some form of a local positioning system is required to provide the system with the user’s location.

GPS-based systems seem to be more appropriate for solving the navigation problem. The combination of a GPS and a GIS system lends itself to position-based navigation (sometimes called pilotage or piloting), as described in [1]. As for the identification problem, GIS systems can identify only larger objects such as buildings.

B. Tag-based Systems

This type of system is operational for both indoors and outdoors applications. Usually, the system has two major components: transmitters (or tags) and receivers. The tags are placed on the objects in the environment that are significant for
the visually impaired person in the environment. To read the signals sent by the tags, the user carries a receiver.

NAVI (Navigational Assistance for the Visually Impaired) is a system that has been developed at the University of Rochester [6]. It uses radio signals to gauge when someone is close to passive transmitters that are deployed throughout the environment. The user’s portable device combines a CD player with a mobile Radio Frequency Identification (RFID) tag reader. Once a tag has been detected, the corresponding track on the CD is triggered to play a spoken message to the user about the detected object.

In the Chatty Environment [7], each object in the environment possesses an electronic beacon which creates a virtual aura around the object. When the user moves into the aura of the object (or when the object such as a bus moves towards the person) the device, carried by the user, informs the user about the object’s existence and offers a standardized interface for interacting with it. The size of the object’s aura reflects the size and/or the importance of the object.

There are two problems with the environment tagging approach. First, for the system to work some infrastructure needs to be in place. In some situations, this infrastructure is prohibitively expensive to build. Secondly, the reliability of the tagging devices themselves (transmitters) are often an issue. They can fail completely or their power expire. Although these issues may be resolved in the future as this technology evolves, the tagging approach may not be feasible to adopt on a large scale in the foreseeable future.

Tag-based systems, as portrayed in the literature, seem to be more suitable for solving the identification problem. They facilitate the identification of small objects, such as grocery items in a super market, as well as large objects such as buildings and vehicles. Because the system typically doesn’t provide any information about the location of the user, these systems provide very little help with the navigation problem.

C. Camera-Based Systems

Inherently, camera-based systems are free of the limitations of GPS-based and Tag-based systems. Cameras can be used indoors and outdoors, and require no infrastructure to be in place to work. They can be used to identify objects in the environment, as well as help with navigation. The usage of stereoscopic cameras in camera-based systems allows the system to measure the distance between a certain object in the frame and the user.

The vOICe is a system that was developed by Dr. Peter Meijer, a senior scientist at Philips Research Laboratories in the Netherlands [8]. The three middle letters represent “Oh I See”. Here, the intent is to translate the images that are captured by a portable camera into highly complex soundscapes which are then transmitted to the user by headphones.

The system consists of a head mounted camera, stereo headphones, and a notebook PC. The system paints a picture of the surroundings of the user by using sounds. Brighter areas sound louder, height is indicated by pitch and a built-in color identifier speaks out colour names.

The Smith-Kettlewell Eye Research Institute is developing a system called Cross Watch [9]. It consists of a a PDA (personal digital assistant) with a digital camera attached to it. The objective is to help visually impaired travelers negotiate traffic intersections.

When finished, the system will be capable of locating a crosswalk (by using the crosswalk stripes) and communicating its location to the user. Another function of Cross Watch is to find the walk light, determine its status (walk or don’t walk), and convey this to the user. Another system that tries to achieve the same objective is being developed by Kyoto Institute of Technology [10]. The system is capable of measuring the length of a pedestrian crossing to within one step length and detect the color of the traffic light [11]. Since the system has only one camera, it does not require camera calibration sophisticated stereo camera systems do.

Tyflos (derived from the Greek translation of visually impaired) is another system designed to convert images, captured by a camera, into verbal messages conveyed to the user [12]. Besides the camera which is mounted on a pair of glasses, the system consists of a laptop computer, a headset, and a microphone. Tyflos offers both object and face recognition capabilities by comparing the objects (or faces) viewed by the camera to those stored in the system’s database.

The biggest hurdle in camera-based systems is the inherent difficulty of analyzing unfamiliar scenes. Even with familiar scenes, other factors such as lighting conditions, camera angle, and the amount of clutter adds to the difficulty of the task.

Like the tag-based systems, the camera-based systems in the literature tend to be geared towards identification applications. These systems that can be applied to the navigation problem are successful only on a relatively small scale (e.g. the navigation of crosswalks).

III. BarKey, An Identification and Navigation System

Here, we propose a new hybrid tag-camera-based system that can be used to solve both the identification and navigation problems. Like Talking Signs, tags are posted on the different objects in the environment. Unlike Talking Signs, the tags are not transmitters; they are barcodes. A portable camera, carried by the user, captures images of the environment, and detects and decodes the barcodes. Each barcode contains the object’s identification number and some information about it. Using the object’s identification number encoded on the barcode, BarKey can access a database to obtain more information about the object.

On the 2D barcode, two major pieces of information about each object are stored: its identification and location. The identification information consists of the type of object (e.g. a store or a bus) and other information that can be useful (e.g., store name, line of business, and hours). The location information is used for navigational purposes. From the location information obtained from different objects, the system can identify the user’s route and navigate the user to the desired destination. Fig. 1 reflects the system’s main functional components.
Since it remedies many of their limitations, BarKey capitalizes on the advantages of tag-based and camera-based systems. Tags facilitate the identification of a predetermined shape and pattern (the barcode), rather than attempting to identify arbitrary objects. The use of a camera to collect the information from the tags (rather than a sensor) eliminates the cost and reliability issues associated with powered transmitters, as well as the need for special purpose receivers.

A. BarKey vs. BarKey SC

There are two suggested versions of the system that cater to the needs of different users and vary in capability and required features. BarKey requires the existence of a GIS database that stores information about the various objects in the environment. Also, BarKey requires a wireless connection to the GIS server. This approach enables BarKey to capitalize on the information storage capabilities of the GIS system, allowing the system to provide richer and more detailed up-to-date information to the user. In addition, navigation information is readily available given the current location of the user, the destination, and the spatial database modeling the environment. The BarKey model is very similar to the GPS-based model described in [1] except that the current user location is detected according to the object(s) nearby, rather than the GPS signal.

However, the power and flexibility BarKey provides comes at a price. Not only is a GIS system required for the identification functionality to work, but also a spatial database is needed for the navigation functionality to work. Although these requirements do not make BarKey any more expensive or difficult to implement than other systems that provide comparable functionality, they still cost time and money. This cost and effort factor can play a role in not implementing the system on a large scale.

BarKey SC (self contained) requires only 2D barcodes to collect information about objects. Although it is dependent on the resolution of the barcode, the 2D barcode is capable of encoding much more information than the typical 1D code can. In many cases, this information is enough for the user to identify the surrounding objects and obtain additional knowledge as well. For example, the name, business hours, and the type of products that a store sells can easily be included on a 2D barcode of a reasonable size.

Although the object’s 2D barcode solves the identification problem, the barcode does little to solve the navigation problem. To remedy this situation, BarKey SC has a database builder to construct its own spatial database of the environment based on the location information BarKey SC extracts from the individual 2D barcode tags. If each tag encodes the coordinates that represent the location of the object, the system learns about the topology of the environment. Gradually, the system builds a spatial database for navigation purposes in the future. An alternative is to provide the user with the spatial database online or through a CD or some other medium.

B. Advantages of Our Approach

Compared to other systems that have been developed to help the visually impaired with the identification and navigation problems, BarKey has definite advantages.

- The system is usable indoors and outdoors.
- BarKey SC requires minimal infrastructure. Only 2D barcodes are required, rendering the system deployment quicker and less expensive than traditional tag-based systems.
- The fact that the system is not dependent on powered tags makes it more robust and eliminates the possibility of failure due to transmitter malfunction or power failure.
- Because the system does not depend on any special purpose transmitters or receivers, the cost of setting up and maintaining the system is expected to be lower than other tag-based alternatives.
- Orientation information can be obtained without the need for an electronic compass, since the location of the tag (to the left or to the right of the user) and its location information determine the user’s orientation.
- The system solves both the identification and navigation problems.
- The system should not have any scalability problems.
- Since the information about the different objects in the environment is stored in the GIS, notes, remarks, and messages can be added by the user so that communities can emerge.
- The size of the barcode can be used in one of two ways. First, it can be used to reflect the size and/or significance of the object tagged. A larger barcode will be recognizable from further distances than a smaller one. Using one size for all barcodes, however, allows the system to estimate the location of the user, relative to a tag based on the size and the perspective distortion of the tag.
- The system can be easily integrated with other camera-based systems.

C. Extracting Information from Tags

BarKey’s functionality is based on the extraction of object information that is encoded in the object’s barcode. Before the barcode can be decoded, it needs to be identified and extracted from the image. According to the camera angle in respect to the barcode, some processing may need to be conducted to rectify any perspective effects. Once the barcode is rectified, it can then be thresholded and decoded. This process is denoted in Fig. 2, and each step will be discussed with application to the sample image in Fig. 4.
In the sample image we use a PDF417 2D barcode to encode 64 characters that represent the name, office number, and office hours for one of the authors. For more information about 2D barcodes in general and PDF417 in particular, refer [13] and [14], respectively. The sample image is shot from a distance of approximately three feet from the barcode at an angle of 45 degrees. The barcode size is 6x2.44 inches and the digital camera resolution is set to 1600x1200.

1) Barcode Detection and Extraction

In this step, the barcode(s) in the image need to be identified and extracted for further processing. Fig. 3 describes this process.

To isolate the barcode, we first apply a Sobel horizontal edge detector to the image which is then dilated using a 3x7 structuring element. To ensure that the barcode has no “holes”, we repeat the dilation process three times. Because the barcode in many environments such as office buildings and hospitals will be surrounded by a non-textured surface, the result of the dilation process should produce a large, white rectangular shaped object, depicting the barcode and other smaller objects. To detect the barcode object, we scan the image with a 40x120 floating window. The window is moved from the top left to the bottom right with no overlap. When the window covers an area of the image of 90% or more white pixels, the barcode is detected.

Once the barcode is detected, the area to crop around is determined as follows:

\[ x_1 = x - 40, \]
\[ y_1 = y - 120, \]
\[ x_2 = x + 40 \times 3, \text{ and} \]
\[ y_2 = y + 120 \times 3. \]

where \((x_1, y_1)\) is the upper left corner and \((x_2, y_2)\) is the lower right corner of the extracted area, and \((x, y)\) is the left upper corner of the floating window. Fig. 4 depicts the different steps.

Figure 2. The process of extracting info from the tags.

Figure 3. The process of detecting and extracting barcodes.

Figure 4. Barcode detection and extraction process
2) Barcode Rectification

To eliminate the perspective effects that the angle of the camera can cause, the extracted barcode is fed to a rectification function. The function takes four points on the original image and the corresponding four points on the rectified image. We employ the four corners of the extracted barcode (see Fig. 5) as the first set of points. The points on the rectified image are determined automatically from the known aspect ratio of the original (non-distorted) barcode.

Based on the notation in Fig. 5, the rectified image coordinates are computed as follows:

\[ x' = x, y' \]
\[ x'' = x', y'' = y, \]
\[ x''' = x'' - \frac{(y'' - y)}{AR}, y''' = y'', \]
\[ x'''' = x''' , y'''' = y'''. \]

Figure 5. The original, extracted, and rectified barcodes.

To identify the four points \((x_1, y_1)\) to \((x_4, y_4)\) automatically, the extracted dilated barcode is processed according to the steps in Fig. 6. After we use a Sobel edge detector to detect the four edges of the barcode (Fig. 7 b.), we run the new image through a Hough transform line detector with a threshold of 40 (Fig. 7 c.) To obtain a single line for each edge, we average all the lines associated with the edge (Fig. 7 d.) Using the detected four lines, we calculate the coordinates for the four corners of the barcode and feed them to the rectification function that has been previously described.

Figure 6. The Identification of the four corners of the barcode for the rectification purposes.

To identify the four points \((x_1, y_1)\) to \((x_4, y_4)\) automatically, the extracted dilated barcode is processed according to the steps in Fig. 6. After we use a Sobel edge detector to detect the four edges of the barcode (Fig. 7 b.), we run the new image through a Hough transform line detector with a threshold of 40 (Fig. 7 c.) To obtain a single line for each edge, we average all the lines associated with the edge (Fig. 7 d.) Using the detected four lines, we calculate the coordinates for the four corners of the barcode and feed them to the rectification function that has been previously described.

Figure 7. The barcode rectification process.

A bar recognition system from Clear Image is employed [15]. The software is capable of recognizing vertical, horizontal, and diagonal barcodes. We set it for accuracy rather than speed and each barcode is recognized (or rejected) in less than 0.05 seconds with a confidence of 95% or more. All the recognized barcodes are correct.

IV. TAG SIZE AND CAMERA DISTANCE

The utility of BarKey depends on the system’s ability to extract and correctly decode the barcodes. There are many factors that can affect the system’s ability to do so. We investigate three factors: the barcode size, the camera distance from the barcode, and the camera angle. Our objective is to see what combination of values for the three factors lead to a successful decoding of the extracted barcodes.

For the experiment, we create five PDF417 2D barcodes of 4x1.6, 5x2, 6.2x4, 7x2.8 and 8x3.2 inches. For each barcode size, we shoot photographs in four different resolutions: 640x480, 1024x768, 1600x1200, and 2048x1536. All the images are photographed from 90, 45, and 22.5 degree angles at a distance of three feet. Tables 1 and 2 summarize the results.

<table>
<thead>
<tr>
<th>TABLE I.</th>
<th>SMALLER RECOGNIZABLE BARCODE SIZE (IN INCHES) PER CAMERA RESOLUTION AND ANGLE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>90°</td>
</tr>
<tr>
<td>2048x1536</td>
<td>4 x 1.6</td>
</tr>
<tr>
<td>1600x1200</td>
<td>5 x 2</td>
</tr>
<tr>
<td>1024x768</td>
<td>8 x 3.2</td>
</tr>
<tr>
<td>640x480</td>
<td>N/A</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE II.</th>
<th>EFFECT OF THE ENCODED DATA SIZE ON THE SMALLEST RECOGNIZABLE BARCODE SIZE (IN INCHES) PER RESOLUTION AT ANGLE 45°.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1024x768</td>
</tr>
<tr>
<td>64</td>
<td>8 x 3.2</td>
</tr>
<tr>
<td>128</td>
<td>9.1 x 4.1</td>
</tr>
<tr>
<td>256</td>
<td>11.4 x 5.6</td>
</tr>
<tr>
<td>512</td>
<td>14.8 x 7.3</td>
</tr>
<tr>
<td>1024</td>
<td>20.5 x 9.4</td>
</tr>
</tbody>
</table>

From Table 1, we can see that the resolution of the camera plays a vital role in determining the size of the barcode. A camera with a resolution of 640x480 or less appears to be of little value for our application, since the barcode must be fairly large to be recognized. Judging by the results in Table 1, we can estimate that a minimum size of approximately 12x5 inches is required if a 640x480 camera is used at a 45° angle. The camera angle has an effect only when the angle is quite narrow. Although there is no difference between the results obtained at 90° and 45° angles, the smallest recognizable barcode size increased by 20% to 25%, when the angle decreased from 45° to 22.5°.

From Table 2, it is evident that the size of the barcode grows at a much slower rate than the size of the data it encodes. This is particularly true for data sizes below 512 bytes. For example, the size of the data doubles from 64 to 128 bytes for an increase of approximately 14% in size, and quadruples to 256 for another 25% increase in size. This is particularly good news for BarKey SC, since a relatively large amount of data...
can be stored on a relatively small barcode. The table indicates that there is almost a perfect correlation between the area of the smallest recognizable barcode and the camera resolution.

V. CONCLUSIONS AND FUTURE RESEARCH

Although many systems for helping the visually impaired with identification and navigation problems have been developed in the last decade, there is a need for a system that is robust, inexpensive to implement, and works for different applications and in different environments. Although researchers claim that their systems currently in use or under development fulfill at least one of these criteria, none of them fulfills them all. Our objective was to meet all the criteria, when we first considered developing an identification and navigation system for the visually impaired.

Although our approach has the potential to render unfamiliar environments more accessible to the visually impaired, there are many issues that need to be resolved along the way. Efficient algorithms must be developed to identify and extract the barcodes and to enhance the images shot in poorly lit environments or by low resolution cameras. Also, user studies are necessary to assess the utility of the system and the most effective setup of the different components (e.g., the location of the tags and the angle of camera).

We believe that our approach opens the door to many possibilities than helping the visually impaired only. Sighted people can also use the system to collect information about the environment, when they are talking to someone or involved in another activity. Those driving on the highway can get information about the attractions available at the next exit or a connection to any database. If the users are connected to a mobile phone or a GPS or a connection to any database. If the users are connected to a mobile phone or a GPS or a connection to any database. If the users are connected to a mobile phone or a GPS or a connection to any database. If the users are connected to a mobile phone or a GPS or a connection to any database.

REFERENCES