34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Safe Trajectory Estimation at a Pedestrian Crossing to Assist Visually Impaired People

Saleh Alghamdi*, Ron van Schyndel, *Member, IEEE*,
and Ibrahim Khalil, *Member, IEEE*

Abstract—**The aim of this paper is to present a service for blind and people with low vision to assist them to cross the street independently. The presented approach provides the user with significant information such as detection of pedestrian crossing signal from any point of view, when the pedestrian crossing signal light is green, the detection of dynamic and fixed obstacles, predictions of the movement of fellow pedestrians and information on objects which may intersect his path. Our approach is based on capturing multiple frames using a depth camera which is attached to a user's headgear. Currently a testbed system is built on a helmet and is connected to a laptop in the user's backpack. In this paper, we discussed efficiency of using Speeded-Up Robust Features (SURF) algorithm for object recognition for purposes of blind people assistance. The system predicts the movement of objects of interest to provide the user with information on the safest path to navigate and information on the surrounding area. Evaluation of this approach on real sequence video frames provides 90% of human detection and more than 80% for recognition of other related objects.**

I. INTRODUCTION

Losing one's vision through accident or illness is a traumatic experience, having both social and emotional impact [15]. One of the most significant rehabilitation steps for people who have recently become blind, or have low vision, is for such people to regain their ability to navigate and react to their immediate environment independently, and therefore to feel that they are in control again [1]. Many contributions have been made over a long period of time to help the visually impaired. Due to the importance of this issue, we have offered a distinct approach based on integration of multiple services to build a model which aims to assist visually impaired people to cross the street independently and safely. Many previous studies have been conducted in the field of object detection [8,17,22], human tracking [18,21], pedestrian detection [16,20] and path planning [19], however the goal of this paper is to describe a component of an integrated system which focuses on the analysis of the properties of objects surrounding a blind user in a cluttered environment. This is part of a larger project involving navigational assistance to blind users.

* Saleh is supported by a scholarship from Taif University.

Saleh, Alghamdi (S3299407@student.rmit.edu.au), Dr Ron van Schyndel (ron.vanschyndel@rmit.edu.au) and Dr Ibrahim Khalil (ibrahim.khalil@rmit.edu.au) are all with the School of Computer Science and Information Technology, RMIT University, Melbourne, Australia.

The paper is structured as follows. Section 2 covers related work, then the overview of the system is given in section 3. Section 4 presents the relevant system components in detail with the results of experiments followed by the conclusion in section 5, the paper concludes with the final findings and suggestions for future work.

II. RELATED WORK

Over the years, many navigational technologies have been developed to aid the blind and people with low vision [2-9]. Much research has been conducted into computer scene-understanding at both a low- and high-level, using many individual modalities such as the navigation tool proposed by Mihajlik [10] which is based on connecting sound generation to a navigation system, using a Digital Signal Processor, and using ultrasonic echolocation which has been applied successfully in mobile robots with a 3D sound generation technique. Another attempt presented in [2] is Electronic Travel Aids (ETAs). An ETA is similar to Mihajlik's device in that both use ultrasonic waves to detect obstacles but ETAs aim to identify objects specifically for visually impaired users. Later research [5] investigated a vision sensor camera to capture images and then process these images and convert them to sound. Almost all this research used a gray level technique to identify objects in images, the 'NAVI' system [5] being a well known example of this approach. Jirawimut et al. [11] presented a visual odometer system using stereo cameras for pedestrian navigation to aid visually impaired people. Indeed, the machine-vision literature is full of techniques and improvements to aid in object identification from images and video [5,7,12,13]. Over the last couple of years, many high-quality projects have been undertaken in the field of robotic path planning [19] and more recent research using Microsoft XBox sensor Kinect to analyze and describe a scene to a blind person [22]. Much research has also been conducted on object identification [5,7,11,12], this is a difficult topic, and our system provides only object position and motion detection, allowing a user to navigate a cluttered but unknown environment of these objects safely. This includes high priority estimation of the distances to close objects, identification of the orientation of some *important* objects and analysis of their movements so as to warn a user of object-user path intersection. In this paper however, we only discuss a specific application of this system involving pedestrian traffic lights and path way assistance.

III. OVERVIEW AND BASIC CONCEPT

The provided system components are Kinect and GPS. Microsoft's Kinect sensor is attached to a helmet, the power is supplied by a battery pack and the Kinect sensor is connected to a laptop for the processing task. We have chosen Kinect because it provides us with two kinds of calibrated images, as shown in Fig. 1, a depth image and a color image. Depth images contain a gray level which represents the distance between objects and the Kinect sensor. Using the Microsoft-supplied API, a linear inverse relationship between the value of pixels as a gray level and the estimated distance exists, where darker objects represent objects which are a greater distance from the camera, and lighter objects represent objects which are closer to the camera. The system can estimate the distance between the camera and the object with high accuracy.

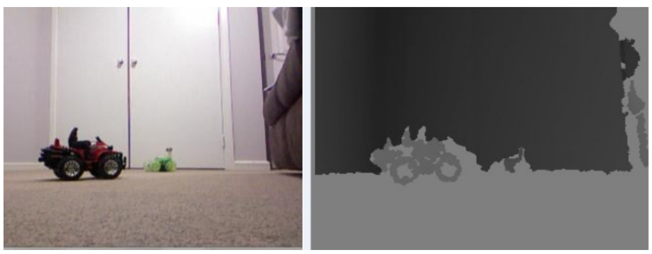

Figure 1. Color and depth image from Kinect

IV. SYSTEM

Our system aims to assist blind people and people with low vision to cross the street safely. It guides a user to the closest pedestrian signal using GPS then guides the user to exact location of pedestrian crossing signal using the normal camera which had been built inside Kinect. We have chosen SURF algorithm [23] because it is invariant of scale and rotation. Fig. 2 summaries the architecture of the provided system. It has been divided into part A as a preparation stage and part B as execution stage. The SURF algorithm requires target images and model images to generate target models which are used to detect target objects in captured images. The final goal of stage A is to establish a list of target models which the system needs to load in order to identify its targets. Since target models can be location-dependent, the system is able to obtain them from cloud-based services.

A. Pedestrian Crossing Signal Detection

After the GPS device guides a user to the pedestrian signal, the system guides the user to the precise location of pedestrian crossing signal platform, ready to cross when the signals allow. To make pedestrian crossing signal detection successful from any presentation angle between 0° and 45°, we have designed a new model which is based on nine targets from different presentation angles. Along with the affine invariance property of the SURF algorithm [23], this provides a degree of pose independence to the recognition process.

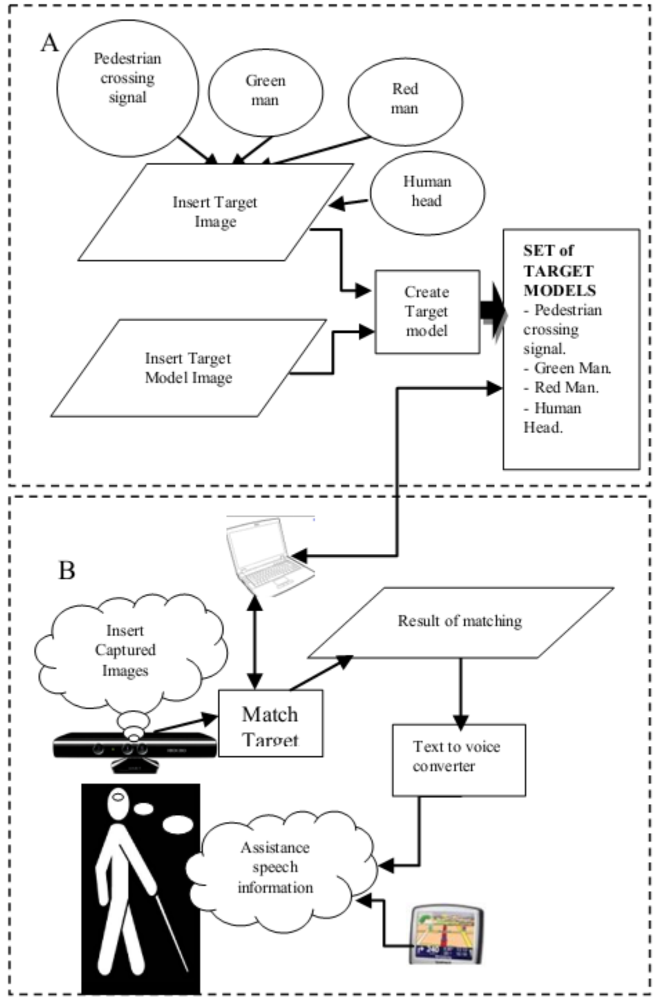

Figure 2. The system architecture

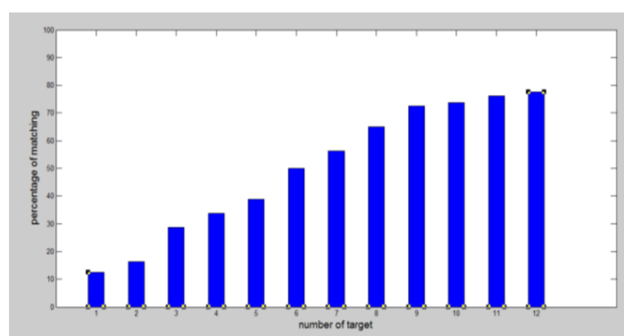

Figure 3. Relationship between number of targets and percentage of successful matching

As shown in Fig. 3 there is a positive relationship between percentage of successful detection of pedestrian signal and the number of target models from one to nine. Therefore, once we apply one hundred random frames on one model we only achieved a positive match in around 12% of cases. In contrast, when we increase the number of used target models to 9, the positive match rate jumped to 72.5%. Because of the nine chosen models cover most presentation angles, the percentage of matching for more than nine models was increasing slightly. A sample of experiments appears at Fig. 4

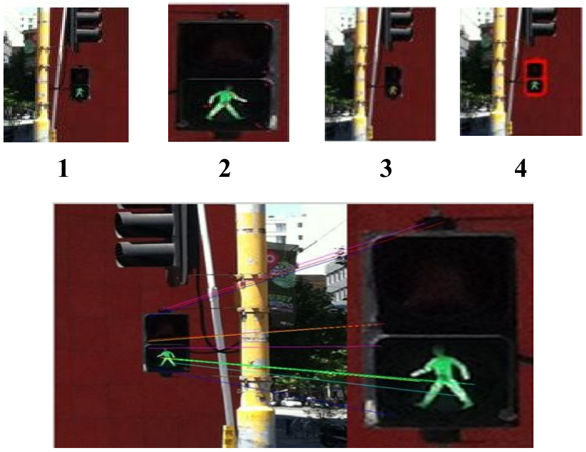

Figure 4. Steps of detection of pedestrian crossing signal

B. Determination of consent to cross street

When a visually impaired person wants to cross the street, he listens for the 'clicker' sound which is emitted by many pedestrian crossing signals, which indicates when it is safe to cross the street, however, sometimes the noise of the city makes it difficult to user to indicate *which* clicker refers to which light, and sometimes they do not work at all. In addition, not all crossings have the clicker facility. Therefore, our system can help the user in these circum-stances. Hence with our system, the clicker sound is secondary evidence of the light state. When a user is facing a pedestrian crossing, he will hear the crossing sound and the system will tell him whether the signal is green or not.

The system can distinguish the green signal from the red signal based on its colour and its template using SURF algorithm, where the green signal shows a man standing with his legs apart, as in Fig. 5a, to represent a person who is walking and the red signal shows a man standing still. as in Fig. 5b. To detect the red man signal, the system removes all color spaces except red then applies the Surf algorithm on that image. If the red man is not detected the same procedure is repeated with the green and green space. The results appear in table 1. In addition, to evaluate the ability of the system to distinguish between signal of green walking man in Fig. 5c and signal of green arrow in Fig 5d, we have performed the green template matching algorithm on two kinds of image. The result is as shown in Fig. 6, a sum of squared differences with signal of man is always less than a sum of squared differences with a green arrow. Less difference means better match. Therefore, we have found the system successfully distinguished green man from green arrow in all 15 random samples.

C. Human Detection

After the system gives the user permission to cross, the system provides the user with information about the surrounding people. We have chosen humans in this case because they are the most common objects at a pedestrian crossing.

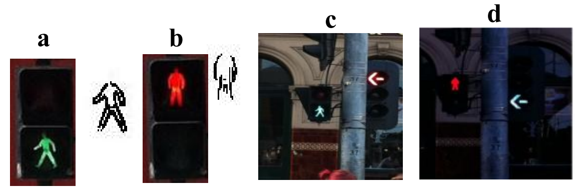

Figure 5. Crossing signals models

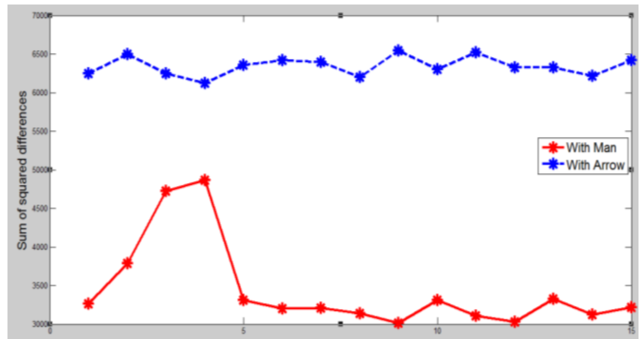

Figure 6. Red line represents the sums of the square of the absolute differences between pixels in the images which contain signal of the green man and the corresponding pixels in the green man model. But the dashes represent the difference degree of Matching between green man model and a green arrow model.

TABLE I. OVERALL DETECTION RESULTS

Detection	Green Man	Red man	Human Head
True Positive	86%	82%	90%
True Negative	95%	91%	-

Using an approach similar to that of Xia *et al* [8], but using SURF algorithm rather than template matching, our system detects humans based on a model of a head. This way a blind person can use other humans starting to cross the road as yet further verification. This is an efficient method because head boundaries are similar in almost all people and they do not differ in varying situations because the head boundary from the front is exactly the same as the head boundary from the back and 'sufficiently' similar from the side. We have got 90% successful detection in 2m range. True negative matches does not make sense in this case.

D. Prediction of Movement of Pedestrians

The objective of this research is to enhance the safety of visually impaired people in cluttered environments, in this case when they are crossing the road. Therefore, when the system detects objects (humans) in front of the user, it will predict the movement of these objects and notify the user when an object is blocking or may block his path. However, if the system notifies the user about all the objects surrounding him, this could be noisy. Therefore, we enable our system to classify objects in front of the user into two kinds of objects, objects which are coming towards the user and objects which are leaving the user.

The system then provides the user with more information on objects which are coming towards him. The classification of objects in front of the user is done by tracking the 3D coordinates of an object – the head location – over time. When the Z distance becomes shorter, from frame to frame, this indicates that the object is approaching and vice versa.

By providing continuous information with unique sound for each nearby object's feature, distance and its movement, we rely on the human to classify and judge for example, people moving with him across the road and people moving the other way.

Fig. 7 shows our testbed system result for three persons and successfully predicts and calssifies their movements. The sound is based on this gradient and on object features.

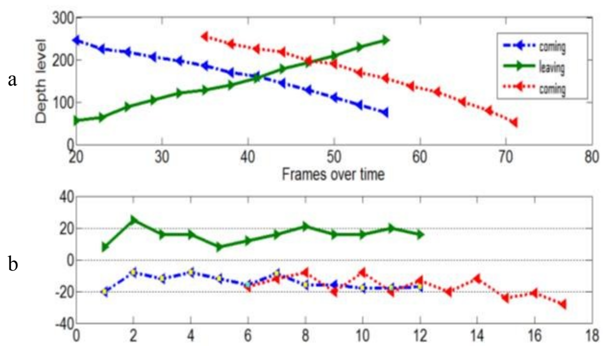

Figure 7. Prediction of movement of surrounding objects (**a.** shows distance of object to observer, **b.** shows its velocity).

V. CONCLUSION

In this paper, we have presented a new system which aims to assist blind and low vision people to cross the street safely. The proposed system detects a pedestrian crossing signal and together with the clicking emitted by the pedestrian crossing, this gives the blind user greater confidence at crossing signals. Furthermore, it detects humans, determines their movements and then notifies the user about humans which may intersect his path. In the near future, we plan to add a pedometer to our system to calculate an approximate velocity of the user to compare it with the velocity of surrounding objects to predict his path more precisely and to get better estimate of intersection points.

VI. REFERENCES

[1] http://www.sciencedaily.com/releases/2007/02/070216221522.htm, accessed 27 March, 2012

[2] G. Balakrishnan, G. Sainarayanan, R. Nagarajan, S. Yaacob, "Stereo Image To Stereo Sound Methods For Vision Based ETA" Computers, Communications & Signal Processing with Special Track on Biomedical Engineering, (CCSP), pp.193-196, 14-16 Nov. 2005.

[3] D. J. Calder, "Travel Aids For The Blind - The Digital Ecosystem Solution", 7th IEEE International Conference on Industrial Informatics (INDIN), pp. 149-154., 2009

[4] M. H. Choudhury, D. Aguerrevere, A. B. Barreto, "A Pocket-PC Based Navigational Aid For Blind Individuals" in IEEE Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems (VECIMS), pp. 43-48, 2004

[5] R. Nagarajan, G. Sainarayanan, S. Yacoob, R. Porle, "An Improved Object Identification For NAVI," in IEEE Region 10 Conference (TENCON), 2004, pp.455-458 Vol. 1.

[6] R. Nagarajan, S. Yaacob, G. Sainarayanan, "Fuzzy Clustering In Vision Recognition Applied In NAVI," Fuzzy Information Processing Society, 2002. Proceedings. NAFIPS. 2002 Annual Meeting of the North American, pp. 261-266, 2002

[7] R. Nagarajan, S. Yaacob, G. Sainarayanan, "Role Of Object Identification In Sonification System For Visually Impaired", TENCON, 2003. Conference on Convergent Technologies for Asia-Pacific Region, pp. 735-739 Vol.2, 15-17 Oct. 2003.

[8] L. Xia, C. Chen and J. K. Aggarwal, "Human Detection Using Depth Information by Kinect", IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 15-22 ,2011,

[9] F. Wong, R. Nagarajan, S. Yaacob, "Application Of Stereovision In A Navigation Aid For Blind People" Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference, vol.2, no., pp. 734- 737 vol.2, 15-18, 2003

[10] P. Mihajlik, M. Guttermuth, K. Seres, P. Tatai, "DSP-based Ultrasonic Navigation Aid For The Blind" in Proceedings of the 18th IEEE Conference on Instrumentation and Measurement Technology, (IMTC), pp. 1535-1540 vol.3, 2001

[11] R. Jirawimut, S. Prakoonwit, F. Cecelja, W. Balachandran, "Visual Odometer For Pedestrian Navigation", IEEE Transactions on Instrumentation and Measurement, vol.52, no.4, pp. 1166-1173, 2003

[12] A. M. Ali and M. J. Nordin, "Indoor Navigation To Support The Blind Person Using Weighted Topological Map" in Int'l Conference on Electrical Engineering and Informatics,. (ICEEI), 2009, pp. 68-72.

[13] G. Balakrishnan, G. Sainarayanan, R. Nagarajan, S. Yaacob, "Stereopsis method for visually impaired to identify obstacles based on distance," Third International Conference on Image and Graphics, pp. 580-583, Dec. 2004

[14] National Federation of the Blind (NFB). 2002. http://www.nfb.org/nfb/default.asp, accessed 27 March 2012.

[15] http://www.mdsupport.org/library/savinglives.html, accessed 27 March, 2012

[16] B. Leibe, K. Schindler, L. van Gool, "Coupled Detection and Trajectory Estimation for Multi-Object Tracking,", IEEE International Conference on Computer Vision,. (ICCV). pp.1-8, Oct. 2007

[17] A. Ess, B. Leibe, L. van Gool, "Depth and Appearance for Mobile Scene Analysis", Computer Vision,. ICCV. pp.1-8, Oct. 2007

[18] M. Rodriguez, S. Ali, T. Kanade, "Tracking In Unstructured Crowded Scenes" Computer Vision, pp.1389-1396, 2009

[19] A. Ess, B. Leibe, K. Schindler, L. van Gool, "Moving Obstacle Detection In Highly Dynamic Scenes", Robotics and Automation, 2009. (ICRA), pp.56-63, May 2009

[20] B. Leibe, E. Seemann, B. Schiele, "Pedestrian Detection In Crowded Scenes", Computer Vision and Pattern Recognition, 2005. pp. 878-885 vol. 1, June 2005

[21] A. Ess, B. Leibe, K. Schindler, L. van Gool, "A Mobile Vision System For Robust Multi-Person Tracking", Computer Vision and Pattern Recognition, pp.1-8, June 2008

[22] D. Bernabei, F. Ganovelli, M. Di Benedetto, M. Dellepiane and R. Scopigno, "A Low-Cost Time-Critical Obstacle Avoidance System for the Visually Impaired.", International Conference On *Indoor Positioning And Indoor Navigation,* pp. 21-23 September 2011.

[23] H. Bay, A. Ess, T. Tuytelaars, L. van Gool, "Speeded-up robust features (SURF)," *ACM Journal of Computer Vision and Image Understanding,* vol. 110, pp. 346-359, 2008.