International Journal of Engineering Applied Sciences and Technology, 2020
Vol. 4, Issue 12, ISSN No. 2455-2143, Pages 715-720

OUTDOOR & INDOOR VISION BASED LOCALIZATION FOR BLIND PEDESTRIAN NAVIGATION ASSISTANCE

K.Tasleem Banu, K.Supriya, K.Sony, M.Chandana, M.Bhavana
Department of ECE
SSE College, Puttaparthi, AP, India

Mr. Baba Fakruddin Ali (M.Tech)
Asst. Professor, Dept. Of ECE
SSE College, Puttaparthi, AP, India

ABSTRACT—In this paper, we present a design of a wearable equipment that helps with the perception of the environment for blind and visually impaired people in indoor and outdoor mobility and navigation. Our prototype can detect and identify traffic situations such as street crossings, traffic lamps, cars, cyclists, other people and low and high obstacles. The detection takes place in real time based on input data of sensors and optical cameras, the mobility of the user is aided with audio signals.

I. INTRODUCTION

Visually impaired people face difficulties in everydays' traffic. A white rod can detect certain obstacles on the ground level in a close proximity, however there are a lot more traffic situations its user should be aware of: quickly moving objects, public transportation, hanging objects, traffic lamps without audio aid, or the speed of other passerbys. Guide dogs are trained for these purposes, but that is not the ideal solution as the training is expensive and time consuming, hence there is a need for a simple, quickly deployable and language independent tool. According to WHO there are about 39 million blind people and another 246 million visually impaired, so the potential market for such a tool is about 3% of the population of the world.

A. Related work

Projects and tools targeting blind people have three main areas. One is the replacement of the white rod with an electronic one. Most of the project focus on navigation, how visually impaired people can get from one point of a city to another. The two most influential projects with this aim are maybe the CASBliP and ARGUS, both founded by EU grants. They do not yet have produced a commercial product, however their experiences need to be considered.

The EU founded CASBlip Cognitive Aid System for Blind People) [1] made an extremely useful survey based on personal interviews with blind people on their most important needs in navigation. The most critical ones are in descending priority the following: poles, holes on the ground, vehicles, stairs, other people, hanging sunshades, traffic lamps, crossing. People, vehicles and holes should be detected from 20 meters away, for the rest a detection distance of 5 meters could be sufficient. They used stereo camera system or infrared sensors for object detection, help the navigation with GPS and accelerometers and use acoustic feedback for signaling.

The ARGUS (Assisting personal guidance system for people with visual impairment) [2] project was founded by the EU as well. This project focused on supporting blind people in autonomous movement and navigation. They use GPS and other radio frequency technologies such as WiFi, RFID and NFC for navigation, and produce feedback on acoustic and tactile interfaces.

An electronic white rod with a mounted camera was developed at the Aarhus University in experimental research project [3]. That tool has a detection range of 2.5 meters.

Several prototypes have been developed for aiding the indoor and outdoor navigation of blind people. The RFID based solution in [4] and a WiFi based solution in [5] have good preliminary result in navigation, however, they are short in range that results in high deployment costs. Client-server architecture based solutions are proposed in [6] and in [7], where the camera and GPS data are sent from the client device mounted on the user to a remote operator who interprets the information and provides audio or tactile feedback to the user. In [8] and [9], [10], the GPS based navigation is extended with a local database in a wearable computer that contains information such as points of interests and obstacles on the current location of the user. A more advanced database based navigation system is proposed in [11], which has an ultrasound imaging based indoor navigation module as well.

Virtual eye based solutions intend to detect and classify object within the detection range of sensors and cameras mounted on the user. Kanna et al. [12] developed an FPGA based device that detects objects based on the information collected by laser and infrared sensors mounted on the leg of the users, and classifies them based on an on- line image catalogue. Rao et al. [13] down sample the image acquired from a camera mounted on the user, and assign a threat level to objects based on their size and position in the image and provide an acoustic feedback. In [14], a low resolution depth map is calculated stereoscopic images, and generate stereo musical sound based the disparity of objects.

B. Our contribution

In the paper, we propose a virtual eye tool that utilizes the experience of the CASBlip projects regarding practical requirement of blind people. Our prototype tool consists of a camera mounted on a helmet, its image is processed with a portable computer carried by the user in real time. Earlier research prototypes were limited to 32x32 images [13] and depth maps of similar size [14], which results in heavy information loss making for instance traffic lamp detection impossible. In our work, we use full scale images, and yet process them automatically in real time. Out image processing algorithms.

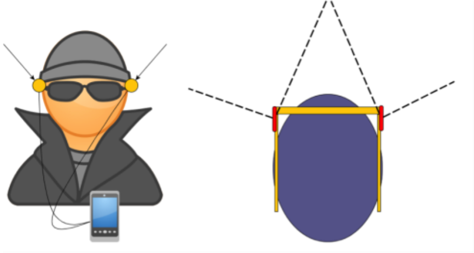

Fig. 1. Image acquisition setup for head mounted cameras

cover the classification of hanging obstacles, road crossings, stairs, traffic lamps, which is much a broader spectrum than in previous virtual eye solutions could provide.

The rest of the paper is organized as follows. In Section II, we give the high level overview of the blind guide prototype and the main tasks of each functional module. We describe the image transformation and object detection algorithms we use in Sections III and IV respectively. Section V describes the threat assessment and the threat to audio signal conversion. Section VI shows the experience of tests and the public demonstration. Finally, we give a brief summary in Section VII.

II. THE OVERVIEW OF THE BLIND GUIDE TOOL

The prototype tool has three main components: a stereo image acquisition module, a data rectification of images and color space transformation take place. compensate for the small radial movement caused by human walk to get a stable image. This compensation is performed before any object detection takes place, otherwise it can not be determined if an object moved in the inertial reference of the camera or the displacement was caused by the small movement of the camera.

For stabilization, we use sparse feature point matching between two successive pairs of frames as shown in Figure 2. After the rectification, we perform a feature point selection in both the left and the right images and execute a feature matching. We perform the processing unit, and a signaling generator unit. In the following, we give a brief overview of these.

Figure 1 shows the schematics of the stereo image acquisition unit. It contains two cameras that are deployed on both sides of the head symmetrically heading forward for supporting peripheral view. The field of view of the two cameras are not required to completely overlap, however dead space must be avoided. Vertically, it is not required to mount the cameras at the same height. In our tool, we fixed the cameras on a helmet. By using a helmet for camera deployment, we can keep hands and feet of the users free, have a better line of sight, and the occlusion caused by hands is not a problem, however, we must compensate for the head movement.

The data connection and the power supply of the cameras are provided by USB cables connected to the Data Processing Unit (DPU). The DPU is a mini PC operated from an external power supply with a capacity enough for two hours' data processing. Functionally, the DPU consists of three modules: a data preprocessing unit, an image processing unit and a data evaluation unit. All of these three units are implemented in pure software.

The data preprocessing unit is responsible for performing necessary tasks that cannot be parallelized. Such tasks are the decoding and the transformation of data extracted from the cameras, and this is the module where the same matching for the next pair of frames, and a matching between the two successive left frames as well. This matching yields a set of motion vectors that represents how objects moved during the last acquisition interval.

Our hypothesis is that most of the feature points are bound to immobile background objects such as poles, trees, houses, therefore the movement of the camera is determined by the most frequent motion vector. We store this vector for the remainder

The image processing unit performs several tasks in parallel: it searches for moving objects in the stereo image, it searches for crosswalks and stairs, and it searches for traffic lamps The data evaluation unit evaluates the traffic situation based on the outputs of the image processing unit. It determines the set of objects relevant with regard to the current trajectory of the user and the object, and it assigns a threat level to each object. The location, size and movement speed objects with a threat level above a threshold are passed to the signaling module.

The signaling generator unit generates an audio sound in the earphones. It must be noted that the volume of the sound is well below the environmental level, but still hearable, so that it does not deprive blind people from their most important sense. The tune identifies the object type, the volume the size or speed and the balance the direction relative to the symmetry line of the field of view of cameras.

III. IMAGE PREPROCESSING

The main tasks of image preprocessing image acquisition, rectification and image stabilization.

A. Calibration

As in any stereo machine vision systems, a stereo calibration has to be performed for each deployment of cameras. The pair of images extracted from the pinhole cameras have an inherent distortion that can be seen near the borders of the captured image. This causes straight lines on the sides and on the top to bend, in our context this means that straight poles can be seen as curves and there are no parallel lines, hence crosswalks and stairs cannot be detected. The distortion depends on the lens characteristics and the deployment of the cameras. For each mount on a helmet, calibration has to be performed once, before the first usage. The output of the stereo calibration procedure is a couple of matrices (camera, distortion, rectification and projection), with which it is possible to define a one-to-one correspondence between the pixels of the two images captured at the same time by the two cameras.

B. Stabilization

The other problem of image acquisition is that and we have to compensate for the small radial movement caused by human walk to get a stable image. This compensation is performed before any object detection takes place, otherwise it can not be determined if an object moved in the inertial reference of the camera or the displacement was caused by the small movement of the camera. For stabilization, we use sparse feature point matching between two successive pairs of frames as shown in Figure 2. After the rectification, we perform a feature point selection in both the left and the right images and execute a feature matching. We perform the same matching for the next pair of frames, and a matching between the two successive left frames as well. This matching yields a set of motion vectors that represents how objects moved during the last acquisition interval. Our hypothesis is that most of the feature points are bound to immobile background objects such as poles, trees, houses, therefore the movement of the camera is determined by the most frequent motion vector. We store this vector for the remainder of the procedure.

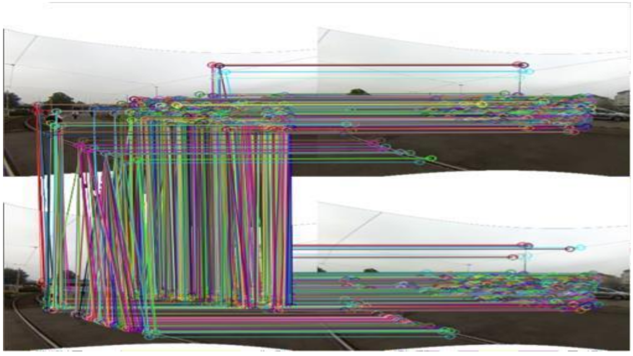

Fig. 2. Sparse feature point matching between two successive pairs of rectified frames for detecting camera movement

IV. IMAGE PROCESSING

A. Moving object detection and classification

In general, calculating the full resolution dense depth map like in [14] is the ideal approach, however computing that for each frame has high demands on the hardware for meeting the real time requirements. The sparse depth map computed from a small set of feature points is only reliable at those point, and the usability obtained depth.

Therefore, we utilize again the motion vector results for object identification, and use dense depth maps. When the motion vectors are determined, the motion vector of the camera calculated in the stabilization procedure is subtracted from all vectors. We apply then a clustering on these motion vectors with locality and texture constraint to determine the set of objects. The locality constraint means that vectors of the similar size are considered to belong to the same object only if they are in the same closure with regard to a maximum distance. The texture constraint allows neighboring points with similar texture as the feature points in the closure to be added to the detected object.

Then, we apply a labeling based on the main direction of the object obtained by fitting a line on the set of points of the object, its size, its movement speed and its position in the image.

If an object has a vertical main direction, slow movement speed and it is vertically in the middle of the image, then that is labelled as a pole. If a similar object has faster movement speed and it is vertically in the middle of the image, then that is labelled as a pole. If a similar object has faster movement speed and the same attributes as before, then that is labelled as a passer-by. The sample is taken from rectangle in the bottom of the left image. The yellow color shows the free path, the decision is made on the right image searched for at the top of the image, and have slow movement speed and horizontal main direction.

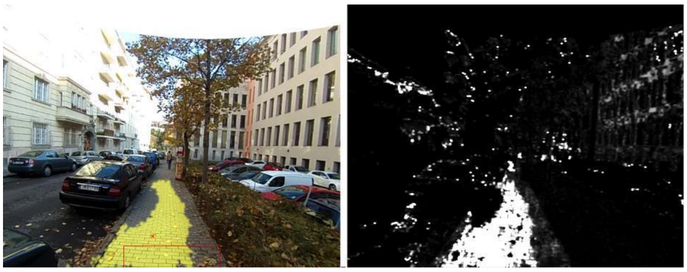

Fig. 3 Detection of free route for the next meters

Figure 3 shows an indoor scenario where we want to pass by a basket. The left image is the original image, where from the middle of the bottom line we start lines radially and check the depth of points at every quarter of the full height. The right image is visualization of obstacles in the way. It shows that there is something in the way at approximately 1 meters away.

Detection of free route for the next meters. The sample is taken from rectangle in the bottom of the left image. The yellow color shows the free path; the decision is made on the right image and there is a free way straight ahead. The basket is classified as a pole, it has slow movement and a vertical main direction.

A problem of object detection, which may arise when the DPU is able to process more than 20 frames in a second, is that the error of matching and the error of depth calculated from the disparity are greater than the movement of an object in that 0.04-0.08 seconds. We handle this problem by matching not direct successor frames, but only every third or fifth. In that time frame, however, the scene may have undergone big changes making matching less effective.

B. Free path detection

The assumption behind our free path detection is that the texture of the free path remains the same in the next meters. Hence, we capture first a sample from a small region in the middle in the bottom of the camera image. Then, we compute the histogram of each color channel for that region, and use a histogram back projection for the whole image. The result is shown in Figure 4, the free pathway is on a cobblestone road, which is difficult to handle by any other texture based technique. In the resulting image, the free passway is highlighted, but not passers, poles or any other obstacles. Moreover, this method is able to detect holes in the asphalt and puddles as those have different color compositions. This method works both indoors and outdoors.

The weakness of this approach is that it may loose track if there is a change of texture of the ground, for instance when the user decides to walk on the grass. Another problem is when the user does not head in the walking direction, however such movement is detected during the stabilization phase, and in such cases free path detection can be switched off temporarily.

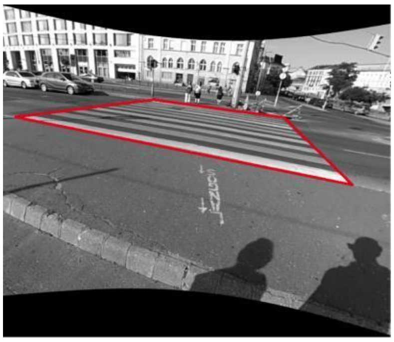

Fig. 4. A detected crosswalk in a distorted image

Ground level estimation is based on the assumption that the larges plane in the image is the ground level itself, which is true for the vast majority of the cases, there are exceptions for instance when the camera is directed on a wall. Our ground detection algorithm selects three non collinear feature points on the free pathway at random, and check how many other feature points from the free pathway are in the same with these points. This procedure is repeated for a predetermined number of times, and the highest number of coplanar points is finally declared as ground level. In general, the number of feature points on the free pathway is very small, those can be found near potholes, puddles and road borders.

B. Crosswalk and stair detection

Crosswalk and stair detection are performed in parallel tasks as these are independent from the optical flow calculations. The algorithms we developed for these two tasks have higher computation requirement, so these are not performed for every frame, these are rerun with the current image of one camera immediately when the last detection is completed.

In crosswalk and stair detection, we utilize that in Europe crosswalks are parallel lanes alternating in white and gray (asphalt color), and after an edge detection these appear as parallel lines, and so do the stairs. Another constraints are that the distance of parallel lines has a very little variance and the length of parallel sections towards the top of the image decreases. We use first a Hough transform for detecting parallel lines in the edge image, and then perform a classification of the candidates using the two former constraints. The output is either a crosswalk or a stair, which can be decided based on the color profile of the bounding box of containing the parallel lines. The one matching the alternating white and gray pattern is a crosswalk, anything else is a stair. Figure 5 show a successfully detected crosswalk. Note that such parallel lines are not necessarily horizontal in the image.

The main problems of crosswalk detection and stair detection are vehicles and passerbys who break the parallel lines. However, the coordinates of detected moving objects are available for this procedure, so it is possible to reevaluate the region after the object has moved away.

D. Traffic lamp detection

Traffic lamp detection is triggered, when the crosswalk and stair detection routine returns with at least one positive hit.

Detecting traffic lamps is far from trivial, the search space is practically the whole image and the lamp itself is very small, it has the size of only a few pixels. As a blind user would be interested in traffic lamps of connected to a crosswalk, we can limit the search space to the region above detected crosswalk candidates. Another input we take into account to further limit the search space are objects detected as poles, because crosswalk lamps are usually attached to them. Finally, we apply color filtering to find the lamps. In Europe, crosswalk lamps are either red or green. The red color filter is particularly effective as it finds only traffic lamps and the red car back lamps, which are immediately dropped because of the region.

V. SIGNALING

The analysis of the traffic situation has three inputs: the set of objects moving in the user’s frame of reference, the set of crosswalk or stair candidates and the set of traffic lamps. The traffic analysis module assigns a threat level to each object detected. The threat level is based on the size, distance and relative speed of the object, and whether that the motion vector of the object is on a collision course with the free path of the user if the object keeps it movement speed and trajectory.

We use the superposition of audio signals to inform users about the threats. Speech generation as an alternative solution is impractical in view of quickly moving object like vehicles, where telling the user about the threat takes more time than the threat persists. The audio signal produced is selected such that it does not suppress environmental noises important for blind people for navigation.

We use six different sound schemes to describe the environment: moving object, stairs up, stairs down, traffic lamp, hole and obstructing subject. The threat level of the object is reflected in the frequency of the sound scheme, an object with a higher threat level have a higher frequency. The distance is expressed with the sound volume, the volume is inversely proportional to the distance, so that a loud sound would make the blind user stop moving. We inform the user about the relative position of the threat with the stereo balance of the sound scheme, the change of balance from one side to the other indicates that the threat crosses the current movement trajectory of the user.

VI. TESTING

During the development of the prototype, we used three types of cameras. For development we used Contour+ wearable cameras with 1280x780 resolution and 25 fps recording speed. These cameras. As Data Processing Unit, processor an 8Gb RAM was uses. cameras have a close API, so in real time tests we used a pair Logitech Quick Cam Pro 9000 USB. We also tested a laptop computer with Intel Core i5 smart glass product with an integrated stereo camera system, a Vuzix 920AR, which however was not able to transfer both images at the same time through the USB connection to the computer.

Fig. 5. A volunteer tries the prototype at the public demonstration

The prototype operates in real time, in one second 25 image pairs are processed with the object detection algorithm, and 1- 3 images with the crosswalk and stair detection procedure in average.

Figure 6 shows a picture where a volunteer tried our proto- type at the Hungarian Innovation TechShow on 30 May 2014. In the demonstration, a blindfold was attached on helmet, anyone could try it. The volunteer was told the meaning of the different sound schemes, and shortly after a brief training he was able to navigate out of the demonstration hall and back, avoiding any collisions.

VII. CONCLUSION

In this paper, we presented our blind guide system, which is designed for helping blind and visually impaired people in their indoor and outdoor movement by providing them real time information about the environment and the traffic situation. The system is based on a stereo camera based vision system and generates audio signals for aiding the navigations. The data acquisition modules are mounted on a wearable helmet, and the processing unit fits in a back bag.

The data processing unit in the current prototype is limited to video inputs. Integrating a mobile data communications module and a SIM card into the DPU allows the system to connect to the services of public transportation companies. The information available through this network communication can be used to help in identifying buses, and thus help visually imparied people to get on the right bus identified by the video processing module.

ACKNOWLEDGMENT

The development of this prototype was supported by Magyar Telekom and Gaia Software Ltd.

VIII. REFERENCES

[1] CASBliP, “Cognitive aid system for blind people,” http://www.casblip. com/, 2009.

[2] A. Consortium, “Assisting personal guidance system for people with visual impairment,” http://www.projectargus.eu/, 2012-2013.

[3] R. Gude, M. Østerby, and S. Soltveit, “Blind navigation and object recognition,” University of Aarhus, Denmark, Tech. Rep., 2008.

[4] B. Ding, H. Yuan, L. Jiang, and X. Zang, “The research on blind navigation system based on RFID,” in International Conference on Wireless Communications, Networking and Mobile Computing, 2007 (WiCom 2007), 21-25 September 2007, pp. 2058–2061.

[5] A. Hub, J. Diepstraten, and T. Ertl, “Augmented indoor modeling for navigation support for the blind,” in CPSN, 2005, pp. 54–62.

[6] O. Gustafson-Pearce, E. Billett, and F. Cecelja, “Tugs – the tactile user guidance system,” Brunel University, Tech. Rep., 2005.

[7] S. Koskinen and A. Virtanen, “Navigation system for the visually impaired based on an information server concept,” http://virtual.vtt.fi/virtual/noppa/mvenue/ navigationsystemforthevisuallyimpairedbasedona ninformation.pdf, 2010.

[8] B. Pressl and M. Wieser, “A computer-based navigation system tailored to the needs of blind people,” in ICCHP 2006, K. M. et al., Ed., vol. LNCS 4061, 2006, pp. 1280–1286.

[9] J. M. Loomis, R. G. Golledge, and R. L. Klatzky, Fundamentals of Wearable Computers and Augmented Reality. Mahwah NJ: Lawrence Erlbaum Associates, 2001, no. 13, ch. GPS-Based Navigation Systems for the Visually Impaired, pp. 429–438.

[10] “Navigation system for the blind: Auditory display

modes and guidance," Presence, vol. 7, no. 2, pp. 193–203, April 1998.

[11] L. Ran, S. Helal, and S. Moore, "Drishti: An integrated indoor/outdoor blind navigation system and service," in Second IEEE International Conference on Pervasive Computing and Communications (PerCom'04), 2004, pp. 23–30.

[12] V. Kanna, P. G. Prasad, S. Amirtharaj, N. Prabhu, and C. Anderson, "Design of a fpga based virtual eye for the blind," in 2nd International Conference on Environmental Science and Technology, vol. 2, 2011, pp. 198–202.

[13] S. K. Rao, A. P. B., A. R. Shetty, Chinmai, R. Bhakthavathsalam, and R. Hegde, "Stereo acoustic perception based on real time video acquisition for navigational assistance," CoRR, vol. abs/1208.1880, 2012.

[14] G. Balakrishnan, G. Sainarayanan, R. Nagarajan, and S. Yaacob, "Wearable real-time stereo vision for the visually impaired," Engineering Letters, vol. 14, no. 2, May 16 2007. CogInfoCom 2014 • 5th IEEE International Conference on Cognitive Infocommunications • November 5-7, 2014, Vietri sul Mare, Italy
347.