A survey on computer vision technology in camera based ETA devices

Amir Ramezani Dooraki

Technology Park Malaysia, Kuala Lumpur, Malaysia
amir.ramezani.my@gmail.com

ARTICLE INFO

ABSTRACT

Abstract—Electronic Travel Aid systems are expected to make impaired persons able to perform their everyday tasks such as finding an object and avoiding obstacles easier. Among ETA devices, Camera Based ETA devices are the new one and with a high potential for helping Visually Impaired Persons. With recent advances in computer science and specially computer vision, Camera Based ETA devices used several computer vision algorithms and techniques such as object recognition and stereo vision in order to help VIP to perform tasks such as reading banknotes, recognizing people and avoiding obstacles. This paper analyses and appraises a number of literatures in this area with focus on stereo vision technique. Finally, after discussing about the methods and techniques used in different literatures, it is concluded that the stereo vision is the best technique for helping VIP in their everyday navigation.

I. Introduction

With recent advances in computer vision, several applications and techniques are designed and developed to help VIPs (Visually Impaired Person) in their everyday life. These attempts can be categorized under two major groups; Sonar-Based ETAs (Electronic Travel Aid) and Camera-Based ETAs. This survey focuses on Camera-based ETAs mainly because computer vision and to be specific stereo vision is the area of this research. Camera-Based ETAs are potentially useful for VIPs in different task such as reading banknotes, clothes labels, detecting and informing about the positions of objects and recognizing obstacles and suggesting safe walk able routes to.

In order to keep this text informative, it is tried to focus only Camera Base ETA devices which are designed to help VIPs in their navigation. In devising ETA devices, in almost all efforts, it is tried to use Sensory Substitution method to orient VIP. In this method, VIP uses another sensory as a substitution for vision sensory. There are two major substitution for this sense; Auditory and Tactile.

Despite of that, in ETA Camera-based devices two methods are basically used to acquire image; Mono vision and Stereo vision. Stereo vision needs more effort because of the necessary processes for generating 3D image and also the need of having two cameras in a normal approach. However, it also has its benefit which is the ability of measuring the depth. As a result, it is possible to measure the distance of different objects in the field of view of stereo rig. Furthermore, this benefit is highly effective in terms of navigation especially because it can detect the obstacles in a more robust way. Thus, in this research because our focus is on the ability of computer vision technique to help VIPs in their navigation, stereo vision is the topic of interest.

The following figure shows the work flow of a typical Camera Based ETA device:

```
┌─────────────────────────┐
│  Image Acquisition Phase │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│   Data Processing Phase  │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│ Person Orientation Phase │
└─────────────────────────┘
```
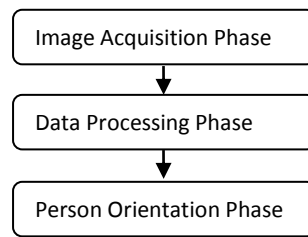
Fig. 1. Work flow diagram of a typical Camera Based ETA device

In the following section some of the most important literatures about stereo vision based ETAs are selected, described and appraised. The selection method is based on the following criteria:

- Innovation of new method.

- Citation count.

- Comprehensive description about a method or technique.

- First and last interesting literatures.

II. Literature Review

In order to make this paper self-descriptive, all literatures are described and appraised under two main groups.

A. Innovative Literatures

1) Stereo Vision-Based Obstacle Detection for Partially Sighted People

In 1997 Stephan Se and Micheal Brady introduced a system useful for VIP to avoid obstacles. This system can be used as a module in another Technological Aid Aimed at helping partially sighted people. In this system "An Stereo Vision-based algorithm (Ground Plane Obstacle Detection) is extended to detect small obstacles for TAPS using RANSAC dynamic recalibration and Kalman filtering."[1]

The system that is described in the paper is part of a more comprehensive system which is called ASMONC (Autonomous System for Mobility, Orientation, Navigation and Communication) expected to provide a full navigation and mobility capability for partially sighted people.

One advantage of this paper is that it is the first of its own that is used Computer Vision technology to help blind people. The method of using disparity map and processing it in order to extract information is a method that is used almost in all other systems which try to help VIPs to avoid obstacles. In addition, this paper talks about Probability and False Alarms which are another advantage for this paper.

However, this paper does not talk about the hardware specification comprehensively (It just mentions of an Ultra-Sparc machine as the processing device). Furthermore, the image acquisition method and also orientation method which indeed are important parts in term of helping VIPs are not described. Nonetheless, the mentioned weaknesses could be a result of the fact that the paper is talking about a component of a comprehensive device and the aim of author is to talk about the other parts like image acquisition and orientation method on other papers but by looking at the paper title it seems that this paper should talk about Image Acquisition and Orientation methods also.

Finally as it is said in the paper "the current DGPR implementation takes 1.5 seconds on the average to process a pair of 128*128 images on an Ultra-Sparc machine. So before it can be actually used by partially sighted, we will need to achieve at least near real-time speed by parallelization and other optimizations."[1] The processing time in the system is not near real-time that is a weakness for the system. However the paper is published at 1997 and the technology is advanced remarkably in compare with that time in terms of processing abilities.

2) A Stereo vision-based aid for the visually impaired

In 1998 a system is introduced by N. Molton, S. Se, J.M. Brady, D. Lee and P. Probert in order to help VIPs to avoid obstacles. This system is in fact part of a more comprehensive system called

ASMONC (Autonomous System for Mobility, Orientation, Navigation and Communication). The aim of this system is to provide a full navigation and mobility capability for partially sighted people.

The system describes that "Sonar is reliable for detecting large and high obstacles, but it is not suitable for detecting small obstacles standing on the ground plane, as the angular resolution insufficient to distinguish between the ground plane and the obstacle." [2] Furthermore, the system's aim is to detect small obstacles, "we are aiming to be capable of detecting obstacles of around 10 cm in height, at a 3-5 m distance, with the vision system".[2]

The paper talks about the hardware specification of the system properly. In addition, the flow of the description regarding the processing procedure of the system is proper. Furthermore, in the last part of the paper the performance of the system is checked in different processors which is a high advantage for this paper in compare with other papers.

In contrary, this paper does not talk about image acquisition and also orientation method which indeed are important especially orientation method is highly important because it is the way that a system transfers its results to its user. Nevertheless, this can be as a result of the fact that this paper is focused only on a component of a comprehensive device. However, the title does not show this.

3) A support system for visually impaired persons to understand three-dimensional visual information using acoustic interface
"This paper outlines the design of a visual support system that provides 3D visual information using 3D visual sound. Three-dimensional information, such as distance map, object recognition, and object tracking required for the visually impaired user, is obtained by analyzing images captured by stereo cameras." [3]

In the system a special type of head phone, bone conduction, is used to generate stereo sound. The bone conduction type is good because it does not block the environment voices and sounds. Furthermore, stereo sound is useful because it can imitate a sound source in a three dimensional space where it is described in the paper that "We have built our acoustic system around the RSS-10 sound space processor made by Roland Corporation. This device enables us to calculate an arbitrary 3D virtual sound space on the basis of HRTFs[1] by just input position, movement vector, and sound source." [3]

The images obtained by the image acquisition part of this system will be processed by a segment-based and correlation-based methods of stereo algorithms according to the necessities of making a 3D image. The operations of measurement, recognition and tracking of 3D objects has been done by analyzing the 3D image with an iteration of segment-based and correlation-based methods. The result is transferred to VIP through a 3D virtual sound system.

The paper has a new approach in helping VIP where it discusses about some different usages of ETA devices such as how to make a VIP able to track a ball while playing games like catch ball. However, it seems the prototype used in the paper needs further development and adaptation in term of transferring the processed data to VIP through its virtual sound system.

Despite of that, the paper mentioned a 3D system with three cameras, but did not mentioned why 3 cameras? Perhaps, two cameras are used as a stereo rig and the third one will be used to increase the accuracy of segmentation, detection and recognition.

4) Real-time pedestrian detection using support vector machines
In the paper a system for pedestrian detection using support vector machine is described. Pedestrian detection is an important task for VIPs which is mentioned in the paper also; "This system is a part of outdoor walking guidance system for the visually impaired, OpenEyes-II that aims to enable the visually impaired to respond naturally to various situations that can happen in unrestricted natural outdoor environments while walking and finally reaching the destination."[4] In term of pedestrian detection, the system should be able to detect obstacles and faces accurately as it is mentioned on the paper that "it detects foreground object on the ground, discriminates pedestrian from other noninterest objects, and extracts candidate regions for face detection and recognition. For effective real-time pedestrian detection, we have developed a method using stereo-based segmentation and the SVM (support vector machine)."[4] This research first separate foreground objects from background then distinguish the pedestrians from the foreground objects. "We used

stereo-based segmentation for object detection, and the SVM technique for pedestrian recognition."[4]

Finally, the system is trained and tested by a data set as large as 528 (140 positive data, 378 negative data are used). The experimental system for test was a standalone machine with Windows XP operating system with a MEGA-D megapixel digital stereo head. The result, as it is mentioned in the paper is that "the system can detect and classify objects over a 320 * 240 pixel stereo image at a frame rate ranging from 5 frames/second to 10 frames/second, depending on the number of objects represented in the image." [4]

The main advantage of this system is that it using machine learning and specifically support vector machines. This will allow the system to learn and recognize shapes as pedestrians. However, 528 samples does not seems to be enough and unfortunately there is no accuracy percentage of the system recognition. All in all, the environment of the testing, the hardware of the system and finally the result of the system are expressed in a proper way.

5) A Smartphone-Based Haptic Vision Substitution System for the Blind
The system described in this paper uses a wireless method in order to transmit the discovered information such as the shape and distance of the objects to a tactile feedback. One of the reasons that this system uses tactile instead of audio feedback, as it is said by author: "Although audio feedback was more articulate than tactile, it hinders the user's ability to hear background noise (i.e. speech, traffic, etc.) and is thus less practical as a travel aid"[5]. Image acquisition is done via a Catadioptric stereo vision system. Furthermore, the images are processed by a smartphone because of its advanced processing capabilities over microcontrollers and because of its portability. The operating system of smartphone is Android and the reason is its wide spread usage among a variety range of devices.

This system produces a disparity map, process it, produce orientation information, compress it and send it to the microprocessor of belt via Bluetooth in order to activate the belt vibrators in necessary times. It is tried to use a small, inexpensive and harmless vibration device. The device is installed in a belt which is wore on the upper back of the visually impaired person mainly because "it is a relatively unused area of the body, providing minimal contraction, and has the somatosensory neuron resolution to accommodate an 8x8 vibrotactile array." [5]

First of all, the system uses a different method in compare with the other papers for generating stereo image. In fact, it uses one camera to make a 3D image. In normal methods, two cameras are used for generating a 3D image. But, in the system described in this article, a Catadioptric device is used instead in order to make the stereo image using the only camera of a typical smartphone. Even though this method is innovative, but a Catadioptric device causes two problems that are: 1-it is bulky, 2-It harder to acquire two proper images, especially when VIP is walking.

Another important aspect of this paper is the use of smartphone as a processing device and Android as its platform. Use of smartphone is an intelligent choice mainly because it is small, convenient and widespread which is certainly an advantage for this paper. Furthermore, using Android as its platform is another wise choice mainly because Android is a widespread platform.

6) A Machine Vision Based Navigation System for the Blind
This paper introduces a computer vision system which uses Stereo Vision technique to recognize planar walk able regions and obstacles. The system basically focuses on finding safe walk able ways and prevents collision with obstacles. It uses two cameras which are installed on the shoulder of VIP in right and left side, a headset and a wearable computer bag. One of the interesting topics in the system is road region extraction which is also an important part in vehicle auto-navigation research. In the paper, homography matrix is used to find corresponding points in the left and right images and as a result generate a stereo image. Another consideration in this paper is the place of the cameras, "Stereo cameras are equipped on the shoulder of the blind, not wearing them on the head as sunglasses, given the width of base line and wobble of head while walking". [6]

All in all, this guiding system do three main jobs; extraction of the road region, detection of the obstacles and measuring their distance. Finally, it transmits the navigation information to the blind person.

This article uses a new approach in compare to other literatures mentioned so far which is recognition of the planar walk able region and extracting it from the stereo image. On the other hand, however, this paper does not give enough information about the orientation method used.

7) Fuzzy matching scheme for stereo vision based electronic travel aid

This paper introduces a wearable system for helping VIPs in their navigation using stereo-based technique. In addition, it uses Fuzzy matching algorithm. "Stereo cameras are employed in the SVETA in order to provide information about orientation, distance, shape and size of the obstacles or objects in front of the user." [7].

Per pair of images acquired by the cameras would convert to a 3D image using Fuzzy matching. As it is said in the paper "the strength of relationship of fuzzified data of the windows in left and the right images of stereo pairs is determined by considering the appropriate fuzzy aggregation operators. However, these measures fail to establish correspondence over the occluded pixels. Left/Right consistence check is performed to overcome these problems."[7] The output of this fuzzy matching which is a disparity map will be converted to a musical tone through Sonification afterward.

The developed prototype in this paper is a helmet molded with stereo cameras, computing device and stereo earphones. The cameras are 1.3 megapixel, progressive scan CMOS with IEEE 1394 interface and low-power. The computing device is a 500MHz processor with 256 MB RAM.

The major advantage of this paper is that it introduced a new method in stereo matching in Camera Based ETA devices; Fuzzy matching. In addition, it has a good explanation of the algorithm. Furthermore, this paper introduces a prototype and explains its hardware properly.

8) A system-prototype representing 3d space via alternative-sensing for visually impaired navigation, 2013 March

This paper studies "the alternative 3-D space sensation"[8] by focusing on moving obstacles as it explains that "Here, we use a scheme for segmenting images and detecting moving objects in the sequence of images"[8]. It uses frame stabilizer, motion detection and motion segmentation using watershed segmentation algorithm. The segmentation happens on depth map which is generated using OpenCV library. Because the device is mounted on a wearable glass which wore by VIP, it can create sudden movements while VIP is moving and this sudden movements will interrupt the efficient recognition of the moving obstacles. Frame stabilizer technique is used here in order to reduce this effect.

In the next step motion information is extracted using spatio-temporal techniques such as motion activity measure, density estimation based on a defined kernel and object detection using watershed algorithm. Furthermore, this information is used for detecting the moving objects and recognize them of stationary objects. Motion Detection happens in parallel with stereo vision, stabilization and obstacle detection. Finally, for transferring the data to VIP, a vibration array (VA) is used and for increasing the accuracy of the orientation and reducing the loss of information in the process of converting the input image ("The effective resolution of the cameras used in our experiments is 640×480 pixels and we convert it into 256×256 pixels for compatibility reasons with the pyramidal approach. This visual information has to be converted into vibrations on a low- resolution array of 4×4 or 32×32 vibrators"[8]) several criteria is used based on the resolution of the vibration array and a high to low pyramidal scheme is used for activating the vibrators.

This paper has some strength points such as having frame stabilization technique for increasing the accuracy and using pyramidal scheme for activating the vibrators accurately. Furthermore, paper explains its different modules and techniques in detail. For example, it has comprehensive description about motion segmentation. Nonetheless, this paper does not explain enough about how it deals with static obstacles. As a matter of fact, the cardinal issue for navigation of VIP is detection of all type of obstacles, static and dynamic, and measuring their distance. Even though, detecting moving objects seems interesting but it does not add extra advantage to an obstacle detection system.

9) Visual Navigation Aid for the Blind in Dynamic Environments

In this paper, recognition of the obstacles in a cluttered dynamic environment using the obtained 6 degree of freedoms using ground plane is discussed. Furthermore, this paper uses semi global block matching for generating depth map and V-Disparity for discovering ground plane in the depth

map. In addition, this paper enhances the v-Disparity method in order to meet their need of dealing with head movements by using of Inertial Measurement Unit (IMU) which gives the information about degrees of freedom. Furthermore, it uses ground plane in order to extract six degrees of freedom and divides it into "1) motion of the ground plane and 2) motion on the plane"[9]. Furthermore, the pose of camera will be obtained from the optical flow of the coplanar 3D points of the ground plane.

One strength point of this paper is detecting the ground plane robustly and using different methods. Furthermore, they discuss about the using of IMU in recognition of ground plane. Finally, they enhance the whole system by adding the motion estimation and removing the unwanted ego-motion movements of the VIP head.

10)   3D glasses as mobility aid for visually impaired people
Random Sample Consensus (RANSAC) is a popular method for finding the outliers which is based on iteration. It can be used in combination with V-disparity for finding straight lines in it where they are considered to be ground plane. For example in [10], authors have used a variant of RANSAC for working in the V-disparity domain as the paper explains that "In order to obtain a robust detection of the ground plane in the v-disparity histogram, we have implemented and tested variants of the RANSAC approach" [10]. The mentioned paper used RANSAC to find straight lines in two dimensional space in order to help VIP by detecting obstacles. Finally, even though V-disparity and RANSAC in this are used for processing the depth map image, but the normal and proved to be the better method is the combination of V-disparity and Hough transform technique together.

Despite of that, FPGA is used in this paper where it is one of the strength points of this paper considering all the image processing calculations need to be done for achieving a real time result. Furthermore, tactile based and auditory based methods are used together to achieve the best result in term of conveying the information to VIP.

11)   Assisting the Visually Impaired: Obstacle Detection and Warning System by Acoustic Feedback
RANSAC also can be used directly for finding ground plane in the 3D-reprojected points of depth map. For example, the work in [11] presents "an obstacle avoidance approach based on stereo vision and a simplistic ground plane estimation algorithm that matches the needs of the visually impaired in their everyday life" [11]. In this paper, it is tried to detect obstacles and find the safe route by searching among the reprojected points of depth map in three dimensional space. The idea is to generate a ground plane randomly and searching for points that are on the generated plane, based on mathematic formula the distance of each point with the generated ground plane will be measured and if the distance was less than a threshold then the point will be considered as a point of the plan, and finally doing this process many times, mainly because RANSAC is an iteration based algorithm, and choose the plane with the maximum number of points as the ground plane. The method used in this paper is simple and effective, nonetheless, it is completely random and takes time. Furthermore, it is highly error prone of the quality of the depth map where the low quality of depth map can highly effect the quality of this algorithm.

Despite of that, a polar grid here is used for assessing the potential obstacles. Furthermore, as it is mentioned in the paper, an acoustic feedback is used for conveying the information to VIP. "Beep sounds with different frequencies and repetition inform the user"[11]. One weakness of using audio based methods for conveying the information is that hey hinder hearing of the environment voice. However, in this paper an audio bone conducting technology is suggested that will solve this issue that is a strength point for this paper.

12)   UV disparity based obstacle detection and pedestrian classification in urban traffic scenarios
The normal methods for discovering the ground plane and as a result recognition of obstacles is by using of either RANSAC or V-disparity where they process the depth map. In [12] however, authors used both U-Disparity and V-Disparity image where V-Disparity image is used to extract road profile and V-disparity is used in order to extract object columns bands. The main difference in this work is the fact of using U-disparity for detecting obstacles where they pedestrians also are obstacles at the first place. In many of other literatures, detecting and organizing of obstacle points will be calculated by removing the discovered points of the ground plane, so all the other points that are not part of the ground plane should be obstacles points.

Despite of that, this paper uses SVM as the machine learning algorithm so its suggested system can be trained in order to be able to classify pedestrians where it is mentioned in the paper "different SVM classifiers are trained considering the relevant features on large pedestrian and non-pedestrian image sets." [12]. Finally the best trained model is used for the algorithm to be able to classify pedestrians.

All in all, using both of the U-Disparity and V-Disparity methods in this paper is one of the strength points for this paper. Furthermore, harvesting the ability of machine learning, SVM, for classifying the pedestrians is another strength point of this literature.

B. Literatures which describe a topic comprehensively

1) Stereopsis Method for Visually Impaired to Identify Obstacles Based on Distance

This paper describes an ETA system which helps VIP to avoid obstacles. In addition, this system gives information about the distance of the obstacles. The hardware consists of a sunglass with two mini cameras fit on that, a laptop for processing data and a stereo earphone to convey the information.

In the first step, after image acquisition, objects in the acquired images will be isolated using image processing methods. Furthermore, the stereo image will be generated based on the isolated objects. According to the paper: "Providing entire environmental information to the blind user often confuse the user in predicting the environment. Information has to be optimized so that only the essential environmental information is made available to the user."[13]

Despite of that, in this system images are acquired in 352 by 288 pixels and colorful, then they converted to 64 by 64 pixel and gray scale images to reduce the computation time in processing data phase. One of the main tasks in this system is to identify objects using edge feature extraction techniques. In addition, more processes are necessary such as linking broken edges and noise removal techniques to perform the object recognition correctly. As a result, the system discovers the objects and use them for generating the stereo image. In the last phase, the depth map is converted to stereo sound using Sonification procedure in NAVI, where amplitude of sound is directly related to the intensity of the image pixels and the frequency of sound is inversely related to vertical orientation of pixels.

In a nutshell, this paper has a good approach on introducing the necessary processes step by step. Furthermore, the orientation part seems to be efficient however it can have some issues with the objects that are in the same vertical level (distance) but different horizontal positions because both of them will produce similar sound.

2) Census-based vision for auditory depth images and speech navigation of visually impaired users

This paper has proposed an ETA system to help VIPs by using of auditory sensory substitution method for orientation method. For generating 3D image a sparse census transform (SCT) method is used. As it is mentioned on the paper, "the proposed algorithm utilizes a sparse census transform (SCT) and color segmentation to obtain an illumination-invariant depth image."[14]

Despite of that, the result of analyzing the disparity map would be conveyed to VIP through DITSM and HLS methods. DITSM is the enhanced ITSM which adds the depth feature to the method and the HLS is a speech based method to orient VIP. According to the paper VIP can perceive the environment easily and without any problem or necessary training using the speech based orientation method. The result of the system based on the paper is: "In good and poor illuminated environments, the performance is 82% and 80% respectively. The performance of our proposed system was not influenced by various lighting. All objects also commented that the systems would be potentially useful."[14]

One advantage of this paper is that it talks comprehensively about orientation method. It discusses about two auditory sensory substitution methods; sound-based and speech-based. In addition, this paper gives important information about the results of the experiments.

Despite of that, another important advantage of this paper is using of Census transform for stereo matching mainly because it is an illumination-invariance method. In addition, the performance of the system in good and poor illuminated environments are measured which are respectively 82% and 80%.

However, the paper does not mention about the hardware specification of the system in testing environment which is a weakness for this paper.

3) A navigation aid for the blind using tactile-visual sensory substitution

In this paper a compact, wearable device with the ability to convert the visual information to tactile signal is described. "This device enable user to perceive distant objects via different sensory modality" [15]

In this system, visual-tactile is preferred to visual-auditory and the reason as authors say is that "Visual-auditory substitution, however, taxes a sensory modality that is already extensively used for communication and localization. Therefore, in this study we have chosen to focus on visual-tactile substitution." [15] Furthermore, the paper mentions even more reasons for choosing the visual-tactile for orientation such as the similarity between skin and retina in their capability to represent information in two dimensions and integrate signals over time and concludes that visual patterns and tactual patterns are functionally interchangeable. In addition, "a number of visual illusions have also been demonstrated for the tactile sense, indicating that perception of spatio-temporal sequence is not exclusively determined at the sensory level but in fact a feature of central nervous system processing."[15]

Despite of that, as it is said in the paper, it is tried to concentrate more on object avoidance rather than object recognition. This paper emphasizes more on transferring the important information to VIP, "Our strategy is to extract the salient features of the visual input in a pre-processing stage, and then provide this information in a spatially relevant way via a sparse tactile array." [15]

The stereo algorithm is written in C programming language and is compiled by Matlab compiler (MEX). Images are acquired in Matlab by Matlab toolbox and also the servo motors are controlled by Matlab. Median filter is used to remove the noise from disparity map. The disparity map then is divided into 14 vertical sections which have overlap on each other. Each part directly can activate one of the motors on the belt. In addition, the activation of the motors is not linear and is up to the distance with the object. Furthermore, only the objects which are in the walking range of VIP are able to activate the motors. All in all, the prototype is able to process 10 frames per second which is good for a normal walking based on what paper says.

One of the advantages of this paper is its comprehensive description about the orientation part which here is by using of visual-tactile. On the contrary, this paper does not talk in deep about the computer vision part which can be considered as a weakness point for the paper. Nonetheless, this also can be as a result of their focus on orientation of VIP.

III. Summary and Conclusion

In this paper a group of papers in computer vision area related to Camera Based ETAs useful for VIP's navigation are discussed and assessed. As it is shown, different aids for VIP navigation are implemented in these literatures such as obstacle avoidance, pedestrian recognition and finding walk able planar way.

Table 1. Percentage of Different methods and techniques used in literatures

Stereo Vision	V-Disparity	Object Detection	RANSAC	Motion Detection	Machine Learning
100%	33%	40%	26%	13%	13%

Even though, different approaches are used for the algorithm mentioned in the discussed literatures such as object detection and motion detection (respectively in 40% and 13% of the literatures as it can be seen in Table 1) however it can be seen that in 100% of them, two images is acquired and used as input for algorithm in order to generate a depth map where it shows the importance of stereo vision technique for navigation. Making a 3D image is a comprehensive and important part where different papers used different ways in order to achieve it and generate a depth map which by using of that, analysis of images for finding obstacles' places and their distances is possible. Nonetheless, there are some difficulties in obtaining a 3D image because of the matching problems which can be as a result of lack of enough texture in the environment or VIP's wobbling

while he is walking. Finally some may argue that Microsoft Kinect is a better option instead the stereo vision camera. Nevertheless, it comes with some issues also such as portability and its range [16]. But the main issue for not considering it here is because it uses a laser technology for generating a depth map and this fact makes it a bit different by a pure computer vision algorithm that uses normal cameras that is in normal method the images are processed in order to generate a depth map but in Microsoft Kinect the depth map image is generated with a different method.

Despite of that, it can be seen that different image processing and computer vision algorithms are used for processing the depth map. The main idea in more than half of them is to find a ground plane. By finding the ground plane it is easier to find safe route and obstacles. Furthermore, as it is discussed in the previous section V-Disparity and RANSAC are two of the popular methods. Where RANSAC is more based on iteration and V-Disparity tries to find the ground plane by converting a depth map to another space and detecting the straight lines on that space. Furthermore, in some papers a combination of them is used. In a nutshell, with attention to the logic of finding the ground plane as the first step and the articles discussed earlier, using of V-Disparity or RANSAC or a combination of them seems to be the best way for starting a navigation aid application for VIPs.

Another area of discussion in this paper is the machine learning aspect of the literatures. With attention to the advancement of machine learning in the last decade and its benefits in terms of recognition, it seems a very good idea to amalgamate stereo vision techniques with machine learning. For example, it can be used to recognize pedestrians and furthermore to recognize their face using face detection algorithm or it can be used to classify obstacles into different categories such as vehicles, trees and stairs.

Last but not the least, is about the orientation method that is used in order to convey the navigation information to VIP. Two main methods are used in the appraised papers; Tactile-Based and Auditory-Based (Alarm based and Speech based). Even though, a visual-tactile (a navigation aid for the blind using tactile-visual sensory substitution) is suggested in some of the paper mainly because they are discussed about its similarity to visual sense. But, visual-auditory method can transfer more data to VIP especially when it is used in a stereo mode. Some papers may argue that it blocks the auditory sense of VIP and hinder his auditory sense to contact with other people but a bone conduction headphone can be used to solve this problem. Thus, visual-auditory method seems a better choice.

References

[1]   S. Se and M. Brady, "Stereo Vision-Based Obstacle Detection for Partially Sighted People," in *Proceedings of the Third Asian Conference on Computer Vision-Volume I - Volume I*, London, UK, UK, 1997, pp. 152–159.

[2]   N. Molton, S. Se, J. M. Brady, D. Lee, and P. Probert, "A stereo vision-based aid for the visually impaired," *Image Vis. Comput.*, vol. 16, no. 4, pp. 251–263, Mar. 1998.

[3]   Y. Kawai and F. Tomita, "A support system for visually impaired persons to understand three-dimensional visual information using acoustic interface," in *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, 2002, vol. 3, pp. 974 – 977 vol.3.

[4]   S. Kang, H. Byun, and S.-W. Lee, "Real-Time Pedestrian Detection Using Support Vector Machines," in *Proceedings of the First International Workshop on Pattern Recognition with Support Vector Machines*, London, UK, UK, 2002, pp. 268–277.

[5]   S. Akhter, J. Mirsalahuddin, F. B. Marquina, S. Islam, and S. Sareen, "A Smartphone-based Haptic Vision Substitution system for the blind," in *Bioengineering Conference (NEBEC), 2011 IEEE 37th Annual Northeast*, 2011, pp. 1 –2.

[6]   S. Wenqin, J. Wei, and C. Jian, "A machine vision based navigation system for the blind," in *Computer Science and Automation Engineering (CSAE), 2011 IEEE International Conference on*, 2011, vol. 3, pp. 81 –85.

[7]   G. Balakrishnan, G. Sainarayanan, R. Nagarajan, and S. Yaacob, "Fuzzy Matching Scheme for Stereo Vision based Electronic Travel Aid," in *TENCON 2005 2005 IEEE Region 10*, 2005, pp. 1 –4.

[8]   N. Bourbakis, S. K. Makrogiannis, and D. Dakopoulos, "A System-Prototype Representing 3D Space via Alternative-Sensing for Visually Impaired Navigation," *IEEE Sensors J.*, vol. 13, no. 7, pp. 2535–2547, Jul. 2013.

[9]   T.-S. Leung and G. Medioni, "Visual Navigation Aid for the Blind in Dynamic Environments," presented at the Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on, 2014, pp. 579–586.

[10] S. Mattoccia and P. Macrı', "3D Glasses as Mobility Aid for Visually Impaired People," in *Computer Vision - ECCV 2014 Workshops*, L. Agapito, M. M. Bronstein, and C. Rother, Eds. Springer International Publishing, 2014, pp. 539–554.

[11] A. Rodríguez, J. J. Yebes, P. F. Alcantarilla, L. M. Bergasa, J. Almazán, and A. Cela, "Assisting the Visually Impaired: Obstacle Detection and Warning System by Acoustic Feedback," *Sensors*, vol. 12, no. 12, pp. 17476–17496, Dec. 2012.

[12] A. Iloie, I. Giosan, and S. Nedevschi, "UV disparity based obstacle detection and pedestrian classification in urban traffic scenarios," in *2014 IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*, 2014, pp. 119–125.

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

[14] S.-C. Pei and Y.-Y. Wang, "Census-based vision for auditory depth images and speech navigation of visually impaired users," *Consum. Electron. IEEE Trans.*, vol. 57, no. 4, pp. 1883 –1890, Nov. 2011.

[15] L. A. Johnson and C. M. Higgins, "A Navigation Aid for the Blind Using Tactile-Visual Sensory Substitution," in *Engineering in Medicine and Biology Society, 2006. EMBS '06. 28th Annual International Conference of the IEEE*, 2006, pp. 6289 –6292.

[16] Microsoft Corporation, "Kinect Sensor," 2012. [Online]. Available: http://msdn.microsoft.com/en-us/library/hh438998.aspx. [Accessed: 06-Feb-2013].