

Stereo disparity estimation algorithm for blind assisting system

Vimal Mohandas · Roy Paily

Received: 19 June 2012 / Accepted: 10 October 2012 / Published online: 23 October 2012
© CSI Publications 2012

Abstract The ability to explore unknown spaces independently, safely and efficiently is one of the challenging aspects of any blind assisting system. Many of the known existing electronic travel aids do not reveal spatial information to the blind. Therefore in this paper the focus is on obtaining the spatial information with the help of image processing techniques. Different disparity estimation algorithms are studied in order to arrive at a scheme specifically suitable for the blind. Real time images are captured by a pair of image sensors and the proposed algorithm will determine the surface conditions such as curbs, stair case and other general obstacles encountered on the navigational path.

Keywords Stereo disparity · Blind assisting system · Depth perception · Correlation

1 Introduction

About 284 million people are visually impaired worldwide; among that 39 million are blind and 245 million have low vision. About 90 % of the world's visually impaired live in developing countries. There is a wide range of navigation systems and tools available for visually impaired individuals. Mobility aids like walking stick and guide dogs are commonly used by the blind even today. With the advances of modern technologies, different types of mobility aids,

which are generally known as electronic travel aid (ETA), are available. The infra red, ultrasonic and laser sensors are used to detect obstacles within few meters ahead. Sonic Pathfinder and Binaural sonar ETA [1] use echoes from transmitted signals to calculate the distance of the object and make alert tones, which are delivered through either earphones or vibrotactile feedback to the user. Navbelt [2] applies mobile robot technologies to assist the blind to detect, to avoid obstacles and to select a preferred travel path. Teletact or Laser cane uses laser sensor for precise measurement up to 10 m. Many of these ETAs are heavy to carry and often distracting the blind person's natural ability to grab the information from natural sounds. Radio-frequency identification (RFID) based location finding and tracking with guidance has only limited application in some existing routes. Global positioning system (GPS) and bluetooth based ETAs provide information about shopping malls, bus stops and traffic signals, but this will not help much in basic navigation. As the capabilities of ETA system increase, so does the complexity of use and therefore many existing ETAs are not well accepted by the visually impaired community. Dimitrios and Nikolaos have presented a comparative survey among portable obstacle detection systems and finally compared each ETA with certain parameters and calculated the performance of each ETA [3].

None of the methodologies discussed above, give spatial information to people who are visually impaired. The ability to explore unknown spaces independently, safely and efficiently is one of the challenging field of all ETAs. We focus on providing the spatial information to a blind person for his/her navigation, using image processing techniques. Volodymyr and coworkers [4] have designed a system where the blind can interrogate the environment by sweeping a standard white cane back and forth. The system

V. Mohandas (✉) · R. Paily
Electronics and Electrical Department, IIT Guwahati, Guwahati,
India
e-mail: vimalmohandas@gmail.com;
vimalmohandas@iitg.ernet.in

R. Paily
e-mail: roypaily@iitg.ernet.in

continuously tracks the cane location and sounds an alert if an obstacle is detected. The cane needs to point in the direction of obstacle and moreover the stereo cameras are rigidly mounted to a wheelchair for this set up to work. Fernandes et al. [5] have developed a system that detects specific circular landmarks in an environment using a Hough Transform method for circle detection.

In this section, the existing methods for obtaining spatial information using image processing techniques are discussed. Perception of depth can be obtained from a pair of stereo images using a technique called stereo disparity. It is one of the popular techniques employed in rendering 3D environments. The stereo disparity method has been applied in a driving assistance system and in robotic vision [6, 7]. Balakrishnan et. al [8] have developed a stereo vision based electronic travel aid (SVETA) for determining distance using an improved area based stereo matching method. However the blind person can recognize only few basic objects with this technique. Our paper extended the same technique to detect any objects encountered along the navigational path. Also the presented work will not be restricted to specific landmark as in the case of Fernandes work [5]. Though not used for the blind navigation perspective, there are different disparity calculation methods available in literature ranging from simple correlation based pixel matching to sophisticated energy minimization methods. Szeliski and Zabih [9] have provided an overview of stereo disparity methods. In order to reduce the uncertainty of the disparity estimation, adaptive window method was proposed [10] by taking into account both intensity and disparity variances. Stereo disparity using dynamic programming [11] is another technique. But, dynamic programming also results in error especially in case of large disparity variation. All these proposed techniques increase the complexity. Another approach known as belief propagation [12] in stereo is available but this technique also consume lot of time and requires many memory elements. Our work is different from the existing methods in the following aspects. Our main focus is providing information to the blind to distinguish between flat and non-flat surfaces. Our focus is on central areas of the image and thus the large computation effort to find out the occlusion points are avoided. We have selected block based correlation method because of its low complexity aspect. The proposed algorithm is highly suitable for a simpler surface reconstruction. However in order to achieve simpler disparity calculations, we have employed a scheme where the height information is required beforehand. Proximity sensors are often employed to provide the distance to the obstacles [13–15]. Since the height information is known in advance, the remaining part of the disparity calculation becomes easier as will be explained in Sect. 3 where the proposed algorithm is presented. Results and discussion is

provided in Sect. 4 and finally the paper is concluded. The final product envisaged from this work will be in the form of a spectacle that a blind person can easily wear.

2 Stereo disparity

Two complimentary metal-oxide-semiconductor (CMOS) image sensors are employed on either side of a spectacle which is specially designed for the blind. Due to the sensor's different positions on the spectacle, such a view creates two slightly different images of the scene. The different perspectives of the same object as seen by two sensors will be relatively displaced and the difference between the images are popularly known as binocular disparity and it provides depth perception of the visual scene [16, 17]. There are two type of stereo disparities known as horizontal and vertical disparities. Horizontal disparity is the horizontal offset of the same image when projected into two cameras. Vertical disparity is the analogous offset in the vertical direction. By rigidly placing the camera in parallel, the vertical disparity can be neglected and in this work we focus only on the horizontal disparity. The magnitude and direction of two disparities can be used for depth estimation. Let L and R are the left and right side image respectively. $I_R(i, j)$ and $I_L(i, j)$ represents the intensity value at i_{th} row and j_{th} column of right image and left image respectively. $D(i, j, \delta_j)$ represents disparity value at i_{th} row and j_{th} column. A cropped window of right image will be correlated with left image window.

$$Disparity(i, j, \delta_j) = \max(\rho(I_R(i, j), I_L(i, j + \delta_j))) \quad (1)$$

Left image window is slided over right image window and it will be continuously checked whether it will be yielding a maximum correlation. The shift (δ_j) corresponding to the maximum correlation gives the disparity measure.

3 Proposed algorithm to estimate spatial information

The flowchart of the algorithm employed to determine the spatial information is shown in Fig. 1. Initially a stereo pair of images is grabbed using two CMOS sensors. The height information (H) between the sensor and the surface is obtained separately using an ultrasonic proximity sensor. The general view of the cameras and object location is illustrated in Fig. 2. The two CMOS image sensors, each having focal length f , separated by a distance B are placed in the horizontal direction such that their optical axes are parallel to each other. The geometry information of the sensors is pivotal for discriminating the flat surface from the obstacles and for estimating the inclination or declination of the surface. Hence, the two sensors are pointed

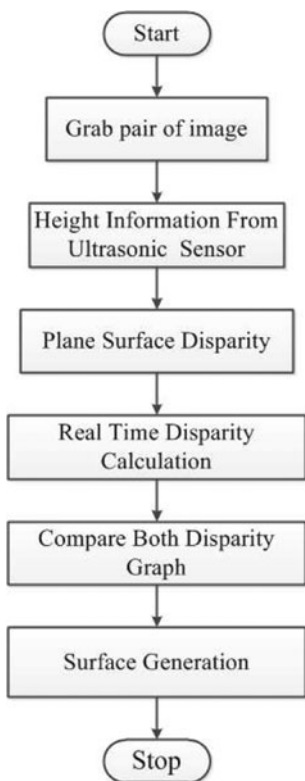


Fig. 1 Flow chart of the proposed algorithm to estimate the spatial information

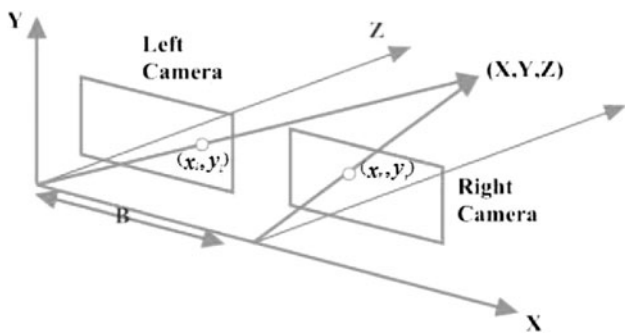


Fig. 2 Camera geometry

downward at a pitch angle of θ with respect to the horizontal plane. As the obtained stereo images are perpendicular to the horizontal direction, the general disparity can be calculated as follows.

$$\text{Left camera } x_l = xf/z \quad y_l = yf/z \tag{2}$$

$$\text{Right camera } x_r = (x - B)f/z \quad y_r = yf/z \tag{3}$$

$$\text{Disparity} = x_l - x_r = Bf/z \tag{4}$$

where B represents the baseline width, f denotes the camera focal length and z denotes the depth. A larger baseline width has pronounced disparity effect and conversely a shorter baseline width would result in almost zero disparity.

However from a practical perspective, the baseline width is about 10 cm which is the end to end separation of a spectacle. In order to find whether the surface is plane or not, the following equation can be used. Assuming that the cameras are at a height of H and the sensors are pointed downward at a pitch angle θ with respect to the horizontal plane, the disparity is given by [7]

$$\text{Disparity} = fB/(H\cos\theta + z\sin\theta) \tag{5}$$

In Eq. (5), except z, all the remaining parameters are constant. Nevertheless, when a person walks, there could be slight variation in θ , however for this work, we have assumed θ to be constant at 45 degree. As can be seen from the equation, the disparity is inversely proportion to the depth (z). Larger the depth, the smaller the disparity will be.

Disparity versus distance graph is plotted in Fig. 3a. The corresponding 2-Dimensional (2-D) representation of the same plot is given in Fig. 3b. The disparity image is coded from 0 to 15, with higher values corresponds to higher disparity. In Fig. 3b red colour denotes maximum disparity and blue colour denote minimum disparity. As the distance increases the disparity decreases. The monotonically decreasing disparity is a representation of a plane surface. This disparity graph is taken as reference for the subsequent calculations. The real time images are acquired using the camera set up and disparity values are calculated. By comparing the real time disparity values with the reference disparity, one would be able to judge whether surface is flat or not. If the surface is plane, then the calculated disparity and reference disparity almost matches. To illustrate the disparity by example, we have considered two ground planes in which one is flat and the other is with an obstacle placed on it. The calculated disparity values in each case are projected on to a vertical plane as shown in Fig. 4. Please note that Figs. 3b and 4a are analogous. The disparity values are higher in the specific location where the obstacle is found.

Camera fixing angle and camera calibration are important factors considered while reconstructing the surface plane. This method can be easily used for surface reconstruction to estimate curbs, steps, slope etc. Also it is accurate enough for identifying even the small objects in the navigational path.

4 Results and discussion

The proposed algorithm is implemented in MATLAB using real time images taken from the laboratory as shown in Fig. 5. The focal length of the image sensors and the height of the two sensors from the surface are 3.3 mm and 140 cm respectively. The size of the captured images is 310×640 pixels. The two

Fig. 3 Disparity plots
a disparity versus range (pixels)
b 2-D representation of disparity in colour

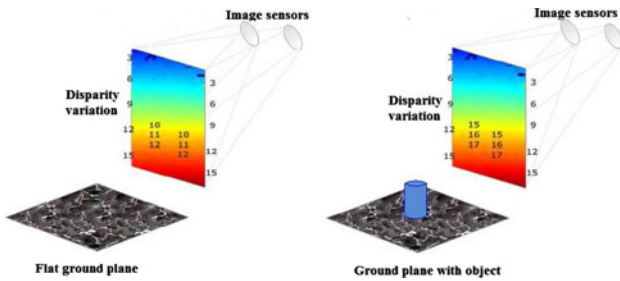
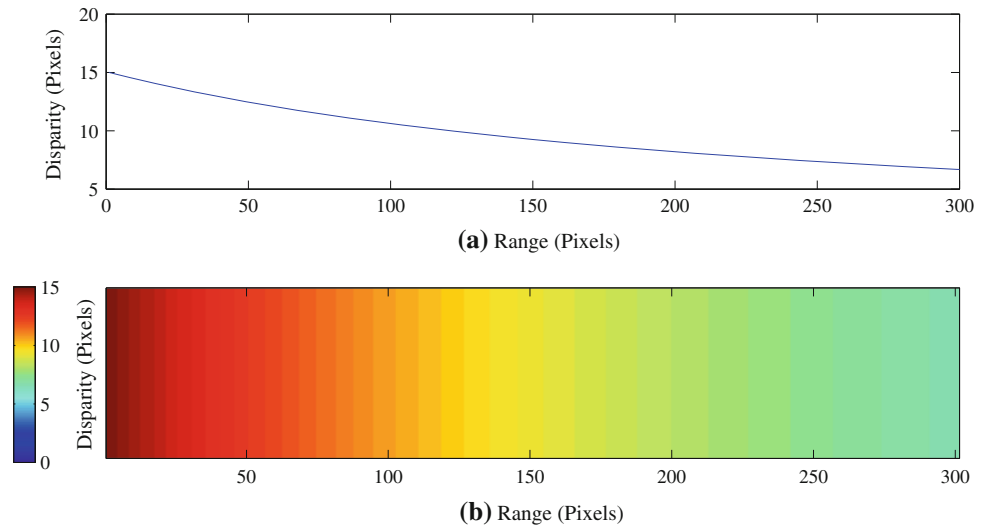


Fig. 4 Demonstration of perspective projection of a surface without **(a)** and with **(b)** an object. The x and y axes of the vertical plane are disparity (in pixels) and range (in pixels) respectively

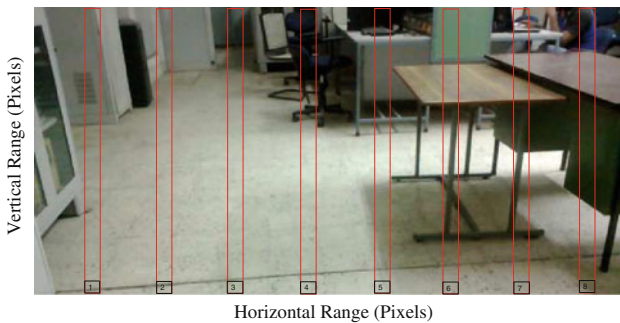


Fig. 5 A test image is segmented into different vertical bands (only eight bands are shown for illustration)

image sensors are focussed towards the ground with an angle of 45 with respect to horizontal plane.

The disparity map of the real time captured images is obtained and it is divided into different bands. For example, the image shown in Fig. 5, is divided into eight bands as indicated by vertical slots of 10 pixels each. For the disparity calculation, a basic block correlation method is employed due to it's lesser complexity. The determination

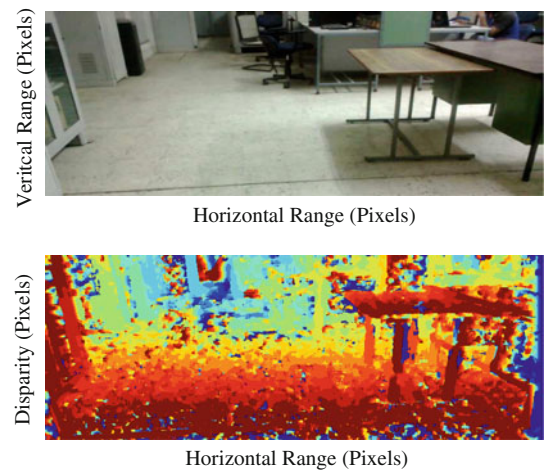
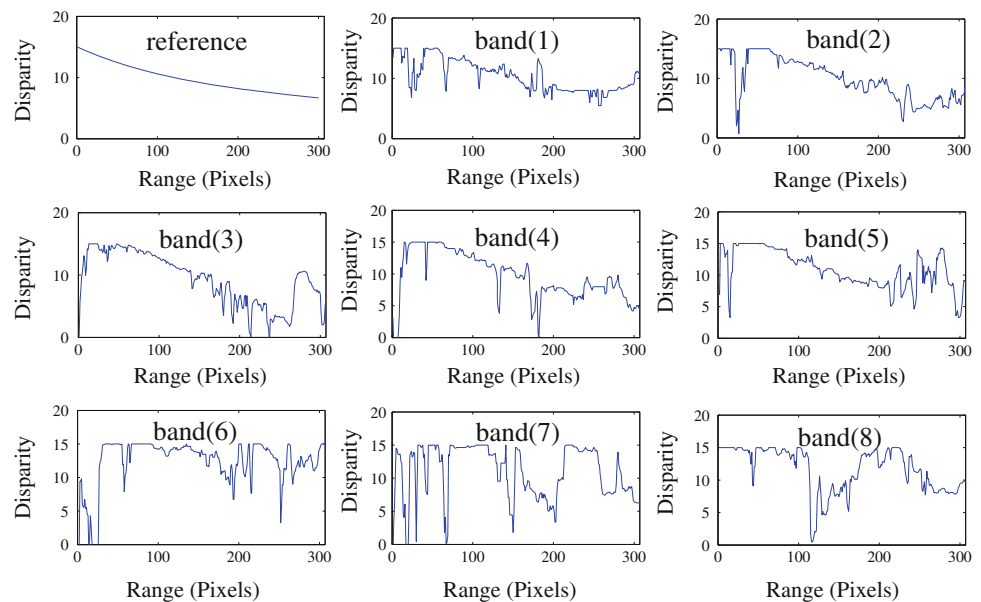


Fig. 6 **a** The right stereo image and **b** it's disparity estimation

of block size of the cropped window will be crucial in this approach. The smaller block sizes give sharp edges but might cause ambiguity about the homogeneous medium. But large block size have good performance in homogeneous medium but is inaccurate at the edges. An optimum block size of 7×7 is considered for the cropped window. The test image and its corresponding disparity map is shown in Fig. 6. The average of the disparity values across the corresponding rows of a particular band was calculated and disparity versus range graph of that concerned band was obtained. Then it is compared with that of the reference disparity graph as shown in Fig. 7. The band without any object should have a disparity graph similar to the reference disparity graph. The presence of an obstacle in a particular zone distorts the disparity graph as illustrated in Fig. 7. Depending on the size and distance of the obstacles the disparity graph will be affected and reveals the spatial information of the image.

Fig. 7 The disparity variation of different bands in the test image



It can be observed from the Fig. 7 that disparity variation of the bands 1–3 are similar to that of the reference graph with little distortion as they do not have any obstacles along its path. But bands 4–5 are having similar disparity variations up to certain distance and beyond which they encounter some obstacles. A closer object will have always a higher shift and correspondingly a larger disparity. Most of the close-by obstacles in the image lies in the bands 6–8 because of which the disparity graph is perturbed a lot compared to the reference disparity graph.

Due to the repetition of the similar patterns correlation based disparity may leads to ambiguous matches. Our proposed approach may fail in certain environments of different lightening conditions as it increases the ambiguity of the stereo correspondence and eventually reduces performance of the system.

5 Conclusions

A disparity estimation algorithm for surface reconstruction was proposed in this paper. A review of the existing ETA for the blind was presented first. Different techniques available for disparity calculation were briefly discussed and a correlation based disparity estimation method was suggested for the navigation of the blind. In order to develop a real time application, we have employed a scheme where the height information is available a priori and thereby simplified the surface estimation calculations. A pair of real images from the laboratory were captured and a block correlation based stereo disparity estimation was employed to predict accurately the obstacles on the navigation path. The system

envisaged will be in the form of a spectacle consisting of a proximity sensor and a pair of CMOS image sensors and it provides spatial information to the blind people about their travelling path through alternative audio or vibration based warning mechanisms.

References

1. Roman K (2002) Binaural sonar electronic travel aid provides vibrotactile cues for landmark, reflector motion and surface texture classification. *IEEE Trans Biomed Eng* 49:1173–1180
2. Shoval S, Ulrich I, Borenstein J (2003) Navbelt and the guide-cane. *Robotics Autom Mag IEEE* 10:9–20
3. Dakopoulos D, Bourbakis NG (2010) Wearable obstacle avoidance electronic travel aids for blind: a survey. *IEEE Trans Syst Man Cyberne C Appl Rev* 40:25–35
4. Ivanchenko V, Coughlan J, Gerrey W, Shen H (2008) Computer vision-based clear path guidance for blind wheelchair users. *Computers and access* 291–292
5. Fernandes H, Costa P, Filipe V, Hadjileontiadis L, Barroso J (2010) Stereo vision in blind navigation assistance. *World Automation Congress (WAC)*, pp 1–6
6. Nedevschi S, Schmidt R, Graf T, Danescu R, Frentiu D, Marita T, Oniga F, Pocol C (2004) 3D lane detection system based on stereovision. *Intelligent Transportation Systems, Proceedings, 2004*, pp 161–166
7. Zhencheng Hu, Lamosa F, Uchimura K (2005) A complete U-V-disparity study for stereovision based 3D driving environment analysis. *3-D digital imaging and modeling*, pp 204–211
8. Balakrishnan G, Sainarayanan G, Nagarajan R, Yaacob Szali (2007) Wearable real-time stereo vision for the visually impaired. *Eng Lett* 14:6–14
9. Szeliski R, Zabih R (2000) An experimental comparison of stereo algorithms. *Vision algorithms: Theory Pract*, pp 1–19
10. Kanade T, Okutomi M (1994) A stereo matching algorithm with an adaptive window: theory and experiment. *IEEE Trans Pattern Anal Mach Intell* 16:920–932

11. Veksler O (2005) Stereo correspondence by dynamic programming on a tree. *Comput Vis Pattern Recognit* 2:20–25
12. Montserrat T, Civit J, Escoda OD, Landabaso JL (2009) Depth estimation based on multiview matching with depth/color segmentation and memory efficient. *Belief propagation*, pp 2353–2356
13. Debnath N, Hailani ZA, Jamaludin S, Aljunid SAK (2001) An electronically guided walking stick for the blind. *Eng Med Biol Soc* 2:1377–1379
14. Ando B, Graziani S, Orazio Lombardo C, Pitrone N (2004) Development of a smart clear path indicator. *Instrumentation and measurement technology Conference* 1:492–497
15. Jin Liu, Jingbo Liu, Luqiang Xu, Weidong Jin (2010) Electronic travel aids for the blind based on sensory substitution. *Computer science and education*, pp 1328–1331
16. Zhen Zhang, Yifei Wang, Dahnoun Naim (2010) A novel algorithm for disparity calculation based on stereo vision. *Education and research Conference*, pp 180–184
17. Ping An, Zhaoyang Zhang, Li Shi (2001) Theory and experiment analysis of disparity for stereoscopic image pairs. *Intelligent multimedia, video and speech processing*, pp 68–71