International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V71I11P225

Volume 71 Issue 11, 236-246, November 2023

Original Article

Assisting Visually Impaired People through Real-time Depth Estimation using Stereo Vision

Moncef Aharchi[1], M'hamed Ait Kbir[2]

[1,2]*STI Doctoral Center, Abdelmalek Essaadi University Morocco.*

[1]*Corresponding Author : maharchi@uae.ac.ma*

Received: 27 July 2023 Revised: 12 October 2023 Accepted: 19 October 2023 Published: 04 November 2023

Abstract *- Visually impaired individuals face significant challenges in their daily lives, particularly during mobility. The loss of vision greatly reduces their ability to detect obstacles, increasing the risks of falls and collisions. However, by employing stereoscopic vision techniques that simulate human binocular vision, it is possible to restore depth perception and help the visually impaired avoid obstacles more effectively. This article presents an in-depth study of the steps involved in creating a depth map using two cameras as part of a system designed to assist the visually impaired during their mobility. The technique utilized by this system relies on the use of stereoscopic vision to provide valuable assistance in detecting obstacles to visually impaired individuals. By analyzing the captured images from the two cameras, the system constructs a depth map that accurately represents the spatial information of the environment. This depth map is a crucial tool in assisting visually impaired individuals in safely navigating their surroundings. The system can detect and alert users to potential obstacles in real time, enhancing their mobility and reducing the risks they face. This article work emphasizes the importance of developing systems that utilize stereoscopic vision to create depth maps for assisting the visually impaired during mobility. By providing valuable assistance in obstacle detection, these innovations have the potential to improve visually impaired people's daily safety and autonomy greatly.*

Keywords *- Camera calibration, Depth map, Rectification, Stereo matching, Visual impairment.*

1. Introduction

People with visual impairments have considerable difficulties moving about and recognizing obstacles. The World Health Organization (WHO) [1] reports that over 2.2 billion individuals worldwide are visually impaired, making it challenging to perform everyday tasks. To navigate independently, individuals with visual impairments often rely on white canes, which help them detect obstacles and sweep the ground for changes in the surface. Outdoor navigation can be particularly challenging as they must carefully listen to auditory cues like traffic and crosswalk signals. Complex indoor environments, like shopping malls or airports, also present additional difficulties. Accessible signage, including tactile markings and auditory cues, is crucial in assisting them. Fortunately, technological advancements offer assistance, including artificial vision devices, mobile applications, and smart glasses specifically designed for visually impaired individuals. However, it is important to note that assistive solutions must be tailored to individual needs, and orientation and mobility training play a vital role in developing visually impaired individuals' navigation and obstacle-detection skills.

Stereoscopy combined with the creation and analysis of a depth map in the field of image processing offers significant benefits for assisting visually impaired individuals in obstacle detection during navigation. Visually impaired individuals often struggle with detecting objects around them due to their limited vision. By enabling depth perception, stereoscopy can play a crucial role in their ability to detect and avoid obstacles accurately.

Advanced image processing techniques make it possible to create a depth map using stereoscopic images captured from different perspectives. This depth map spatially represents objects and obstacles in the visually impaired individual's environment, providing precise information about their position in three-dimensional space. By analysing this depth map, assistive systems can detect depth variations that indicate the presence of potential obstacles. They can then alert the visually impaired individual appropriately, for example, through auditory signals or vibrations, to help them avoid collisions.

Using stereoscopy and depth maps offers several advantages for obstacle detection during navigation. It allows for a more accurate perception of distance and proximity to objects, aiding visually impaired individuals in quickly assessing and reacting to obstacles in their path. This enhances

their confidence and autonomy, reducing the risks of falls or injuries. It is important to note that stereoscopy and depth map technology in image processing do not replace other assistive devices used by visually impaired individuals, such as white canes or auditory aids. However, they serve as a valuable complement that improves spatial perception and obstacle detection during navigation.

In recent years, various researchers have proposed groundbreaking solutions to assist visually impaired individuals in detecting obstacles and achieving independent mobility. One such system, presented in [2], proposes two solutions for visually impaired individuals: smart glasses and smart shoes. These devices integrate various sensors with Raspberry Pi technology to assist navigation and daily activities. The smart glasses may provide features like obstacle detection, object recognition, and audible instructions to help users better understand their environment. The smart shoes, on the other hand, may offer feedback on terrain and obstacles via vibrations or auditory cues. The paper seeks to provide a holistic solution to improve the mobility and independence of individuals with visual impairments by integrating both devices and analyzing the data they generate.

Another remarkable project [2] introduces an innovative computer vision-driven perception system to facilitate the independent navigation of individuals with visual impairments. This system boasts a range of prominent attributes. To begin with, it empowers real-time identification and acknowledgement of obstacles and mobile entities in potentially congested urban settings, all without the need for pre-existing knowledge about these objects. Additionally, the system includes a building and landmark recognition component to improve navigation and positioning in urban areas where GPS can be unreliable. For indoor use, the system can learn and identify user-defined objects of interest and provide audio feedback to guide users in locating these objects. The system conveys feedback through bone-conduction headphones, enabling users to receive system alerts without obstructing ambient sounds. Notably, the entire system is seamlessly integrated into an Android smartphone, ensuring it is easy to carry, unobtrusive, and cost-effective. This technology promises to greatly improve the independence and safety of visually impaired individuals across various environments.

Additionally, another article [4] describes a prototype for a navigational aid device that is intended to help visually impaired people navigate both indoor and outdoor spaces. The device in question is known as the "smart white cane," and it is intended to be affordable. The device combines smart sensors with GPS and GSM technologies to aid users in various aspects of navigation, including obstacle detection, depth perception, water detection, and location finding. Notably, including a GSM module allows users to send SOS messages to family members during critical situations. The smart white cane offers a pragmatic, cost-effective, and user-friendly approach to improving the mobility and safety of individuals who are blind or visually impaired. In a broader context, these state-of-the-art solutions show significant potential in enhancing the independence and mobility of individuals with visual impairments. By harnessing advanced technologies like SONAR, IR, RFID, Bluetooth, and voice recognition, these systems deliver real-time obstacle detection, facilitate straightforward navigation in public transportation, and provide personalized assistance customized to the unique requirements of each user. As research and development in assistive technologies continue to progress, visually impaired individuals can look forward to a future with improved accessibility and inclusivity, empowering them to lead more confident, fulfilling, and independent lives.

A noteworthy project [4] presents a novel method for visual rehabilitation using stereo-vision technology. The device helps visually impaired people navigate by combining stereo vision, image processing, and sonification. It is integrated into a helmet. By picking up on obstructions and delivering depth information, stereo cameras pick up visual information from their surroundings and process it to bring out the most important features of the picture. This method simulates human vision. The visually impaired can sense their surroundings thanks to the processing of the data and its delivery as musical stereo sound. Experiments conducted both indoors and outdoors have proven how successful this image-processing technique is at identifying objects and providing navigational support.

Despite significant advancements in computer vision and image processing, there remains a critical gap in our ability to provide accurate and reliable depth perception for various applications, including autonomous navigation and object recognition. While effective in many respects, the existing camera model has demonstrated limitations that hinder its performance, especially in scenarios with varying lighting conditions and complex geometries. These limitations have led to a critical need for a more robust and accurate camera model to overcome these challenges and contribute to the broader field of computer vision.

This study addresses this pressing issue by exploring implementing an alternative camera model. Stereoscopy, combined with the creation and analysis of depth maps in the field of image processing, offers significant benefits for assisting visually impaired individuals in obstacle detection during navigation. Visual impairment often limits their ability to detect objects around them, making stereoscopy's depth perception capabilities invaluable. Advanced image processing techniques can create depth maps using stereoscopic images captured from different perspectives. These depth maps spatially represent objects and obstacles, providing precise three-dimensional spatial information.

Analysing this depth map allows assistive systems to detect depth variations that indicate the presence of potential obstacles. This enables the system to alert the visually impaired individual, typically through auditory signals or vibrations, helping them avoid collisions.

Stereoscopy and depth maps offer several advantages for obstacle detection during navigation. They enhance spatial perception, providing more accurate information about distance and proximity to objects. This boost in confidence and autonomy reduces the risk of falls and injuries. It is important to note that stereoscopy and depth map technology complement other assistive devices used by visually impaired individuals, such as white canes and auditory aids. They do not replace these tools but serve as a valuable addition that enhances spatial perception and obstacle detection during navigation. Depth estimation plays a crucial role in numerous machine vision applications, particularly in scenarios involving robots and rovers that need to perceive their surroundings [6].

However, directly determining the depth of a pixel, which represents the distance between an object and the camera, is not a straightforward process. To illustrate, one can observe a notable disparity when viewing an object with one eye and subsequently switching to the other; the perspectives vary slightly. This difference is referred to as disparity and is directly linked to the object's depth [7]. Leveraging this principle, stereoscopic cameras incorporate two lenses that capture images from different angles, facilitating the computation of pixel depth in an image.

By analysing the disparities between corresponding pixels in the two images, these cameras can provide valuable depth information, enabling accurate spatial perception in various machine vision tasks. There are simple and fast algorithms that can generate depth maps from stereo images with very high resolution [7]. The cameras must be calibrated to calculate the depth from images taken through a stereoscopic camera. This is a very important step that can be done using test images such as the checkerboard, for example. The two cameras are used to capture the checkerboard in different poses. The technique of estimating the camera characteristics from these pictures is known as camera calibration [9].

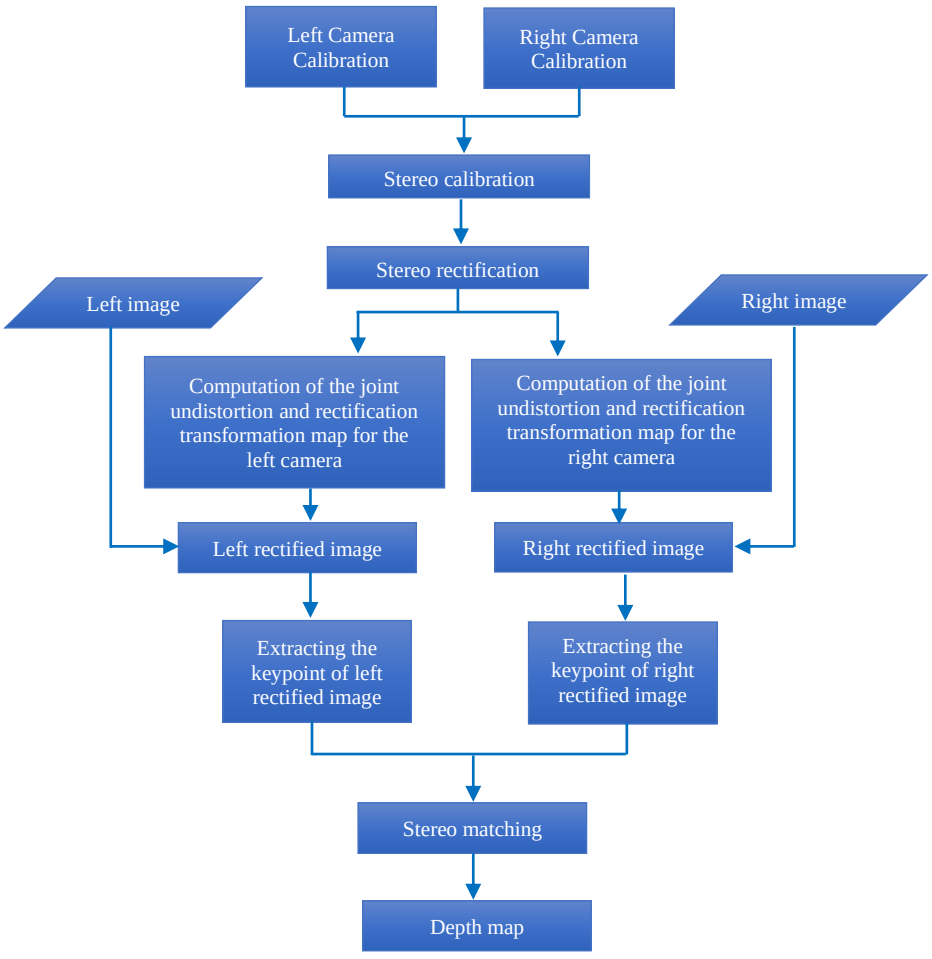

Fig. 1 3D reconstruction process [12]

An important phase of in-depth estimation is the rectification procedure. It addresses the small variations, or disparities, between corresponding points in two images taken by the two cameras [9]. The disparity is directly proportional to the objects' depth in the scene. It is necessary to correct the two images in order to compute this discrepancy precisely. Rectification involves aligning the two images so that corresponding points lie on the same scanline, essentially transforming the images into a common plane.

This alignment is typically done to simplify the disparity calculation, specifically in the depth direction. As a result, the disparity map obtained from the rectified images provides a one-dimensional representation of the depth information.

Alternatively, the cameras can be perfectly aligned before capturing the images, eliminating the need for rectification. In such a setup, the cameras are positioned with precise calibration, ensuring that their optical axes are parallel and have no rotational differences. This arrangement allows for straightforward disparity calculations and depth estimation directly from the original images.

The acquisition of calibrated and rectified images is the first step in the depth recovery system. For applications that require quantitative measurements, camera calibration is extremely important [11].

Three steps make up the 3D reconstruction method [13] :

- Camera Calibration/Rectification: The cameras utilized in this process must be calibrated first. Camera calibration can be used to get both internal and external camera parameters. These options are required to correct and undistort the images such that the epi-polar lines are aligned along the same axis,
- Correspondence: This phase involves computing the difference between each and every point in the corrected images. By limiting the discrepancy to the X-axis, the process is made simpler,
- Reconstruction: The difference is proportional to the depth. The disparity map can be used to calculate the depth map. The depth map makes it simple to understand depth.

2. Materials and Methods

2.1. The Stereo Camera Setup

In a stereo camera setup, two identical cameras are normally mounted at a set distance apart. In the industry, standard stereo camera installations use an identical pair of cameras [14].

The most critical thing is to have the cameras parallel and tightly attached. This stereo camera system consists of two USB cameras (of the same kind) mounted on a rigid base.

Fig. 2 An example of how the two cameras might be used to navigate inside a house

This navigation assistive system uses two stereo cameras placed on the visually impaired person's glasses; it employs stereoscopic logic to create a depth map and detect nearby obstacles. It is important to emphasize that these two cameras are just a part of a larger system that the visually impaired individual will carry. While this technology represents a significant step forward in enhancing the navigation experience for the visually impaired by detecting nearby obstacles, it is crucial to acknowledge that its use cannot guarantee 100% autonomy in mobility. The continued use of navigational canes or guide dogs will always remain essential for complementing the system's capabilities.

The cameras used have the following parameters:

- Resolution: 1280x720 pixels
- Frame rate: 30 frames per second
- Viewing angle: 140°
- Dimensions: 23x23x23 mm
- Weight: 15 grams

Fig. 3 The stereo camera setup

2.2. The Importance of Calibration and Rectification

The research attempted to generate a disparity map using photographs obtained from their stereo setup without performing any calibration or stereo rectification to demonstrate the necessity of stereo calibration and stereo rectification. The disparity map produced by the stereo camera setup that has not been calibrated is extremely noisy and useless. To make the point correspondence search easier, the relevant key points should have the same Y coordinates [15]. In Figure 4, It can be observed that the matching lines between some corresponding spots are not perfectly horizontal.

Fig. 4 The Disparity map

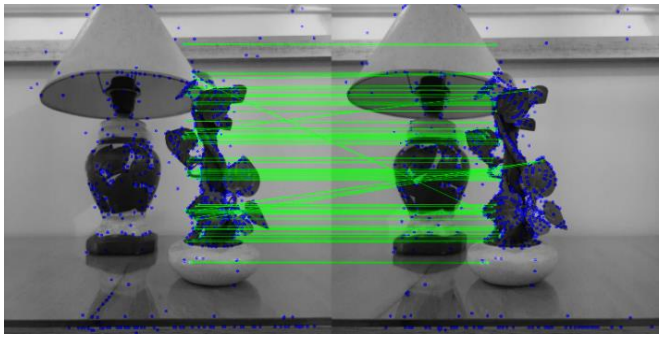

Fig. 5 Matches between corresponding points

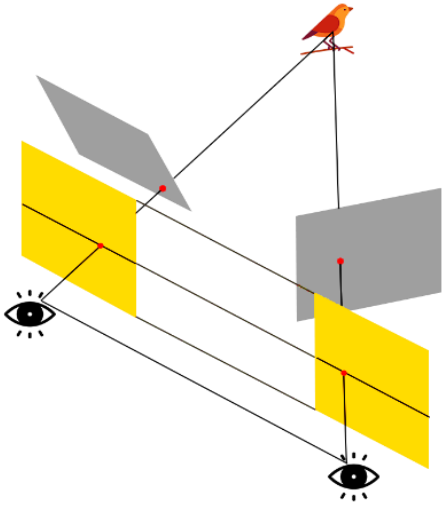

Fig. 6 Process of stereo rectification

A pair of stereo pictures with point correspondence are shown in Figure 5. Blue dots show the key points, while the green lines represent the correspondence between them. The disparity map is now less noisy than the previous one. In this case, the Y coordinates of the respective key points are the same.

This result is only possible with parallel cameras. This is a sort of two-view geometry in which the images are parallel and are only connected horizontally. This is important since the disparity map is created using a method that only searches for point correspondence horizontally [16].

Rather than physically adjusting the cameras to obtain a good disparity map, the approach of Stereo Image Rectification is employed. [17]. The technique of stereo rectification is illustrated in Figure 6. The objective is to project the two images on a plane paralleling the optical center line. This ensures that the matching points have the same Y coordinate and are only connected horizontally.

2.3. Steps for Stereo Calibration and Rectification

Calculating the disparity between each pixel in an image is made simpler by image rectification. It appears that simply a horizontal shift was used to capture the two images acquired by the stereo camera. After this rectification step, the epipolar lines on the images become parallel to the horizontal axis [18].

Performing Stereo rectification requires two important tasks to be performed [19]:

- Detecting key points in each image,
- Find the best key points that match the two images to calculate the projection matrices.

2.3.1. Detecting Keypoints

Usually, both cameras need to be calibrated first before starting. In this case, the two cameras are calibrated and have the same characteristics (same model). The conventional SIFT algorithm will be used [20].

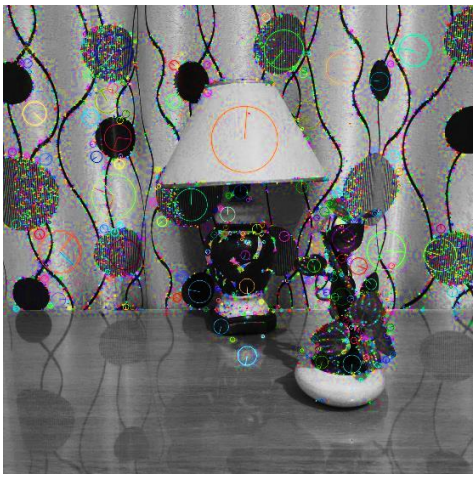

Fig. 7 Detecting key points using the SIFT algorithm

2.3.2. Matching Keypoints

After determining the key points for each image. There will be differences in the points detected because they are taken from a slightly different angle. As a result, to do a stereo rectification, the initial step involves matching the key points between both images. This enables the detection of points which are present in both images, as well as their positional differences.

The matcher FLANN is one of the most effective algorithms for matching key points between them [21]. Using the K-nearest-neighbor search, the FLANN matcher can find the best possible matches between each 2 similar key points on the two images based on their distance.

A limitation of this method is that it generates an excessive number of matches, which can impede progress to the subsequent steps. To mitigate this issue and retain only high-quality matches, the selection process is focused on choosing the best matches, as described by Lowe [21]

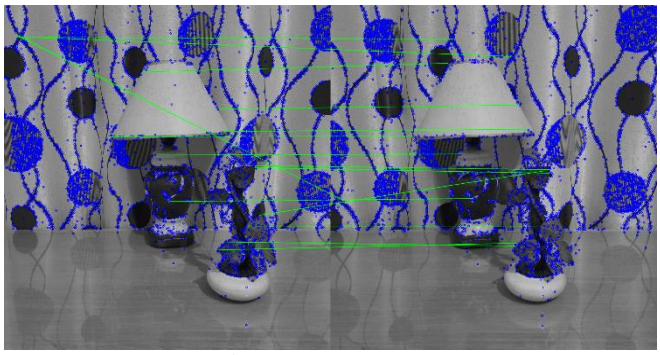

Fig. 8 Visualization of some best-matched key points between two images

2.3.3. Stereo Rectification

Up to this point, the process has involved establishing correspondences between the two images within a 2D plane. The images must now be arranged in a 3D relationship. To achieve this, the application of epipolar geometry is employed.

The fundamental matrix can be used to project points of the scene in the two images captured in the same scene. It is a relation between two images in the same scene [23].

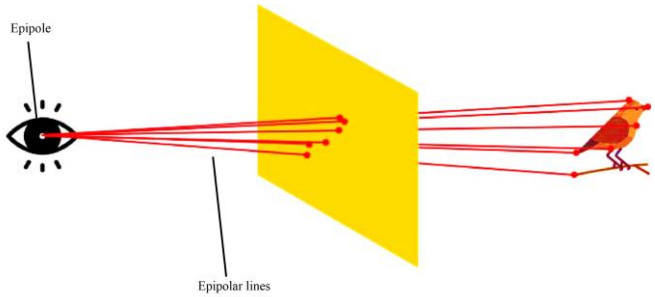

Fig. 9 Basics of epipolar geometry

2.3.4. Epipolar Constraint

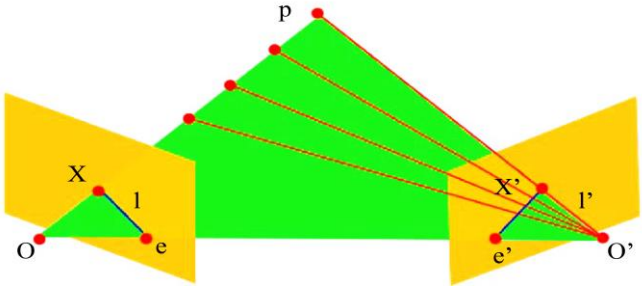

Fig. 10 Epipolar constraint

The two yellow rectangles represent the image planes of the two cameras. The point P represents a point of a scene which is present in the 2 images. The projection of the point p on the 2 planes is x and x'. However, they will not be in the same position:

The main objective is to calculate the depth by determining the disparity between the two images. The objects closer to the camera have a higher offset than those farther away; as can be seen when comparing the two images, they are almost in the same position on the two images [23].

Fig. 11 Matching a point from the left image with the right image

So, how to find the point x' in the second image? The goal is to find the correspondence for the key points and any pixel/block in the image to obtain a dense disparity map.

The determination of the correspondence of point x' in the picture o' to lie on the epipolar line is made possible through the utilization of the reprojection matrix between the two images. Without employing this technique, the process of finding a suitable match would entail searching across the entire image, making it considerably more challenging [25].

2.3.5. Epipolar Lines

The epipolar lines, also known as epilines, will be at an angle in the two images. In terms of calculation, this makes the match more difficult. With stereo rectification, pictures are reprojected onto a shared plane parallel to the line connecting the camera centers.

Now that the images are rectified, the disparity calculation becomes simpler since it is enough to make a search all along the epipolar line horizontally on the other image [26].

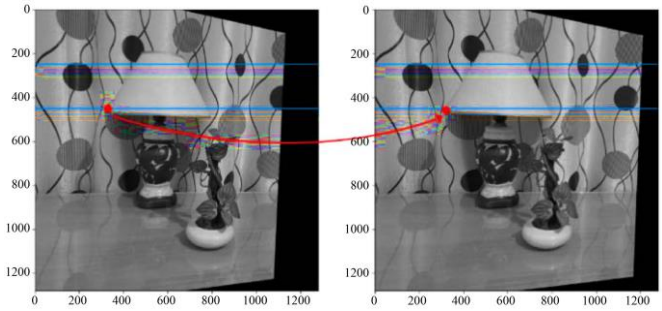

Fig. 12 Epipolar constraint reduces the search to a single line

With the images now rectified, matching between the two images becomes a straightforward process. It simply involves block comparison, and the best match is determined by selecting the block with the lowest cost.

Typically, the normalized cross-correlation (NCC) or the sum of absolute differences is used as the algorithm (SAD).

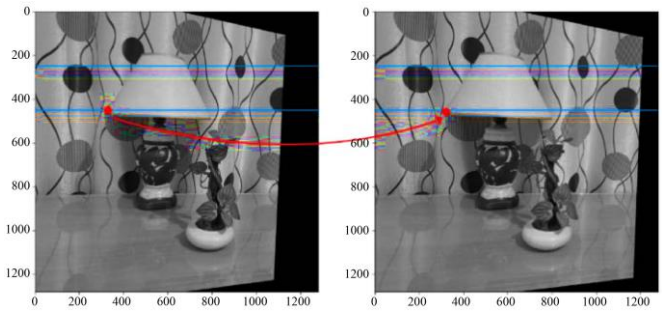

Fig. 13 Stereo block matching

2.3.6. Fundamental Matrix

Given a point in the first image, the epipolar line in the other image can be found using the fundamental matrix [28]. A more generalized approach to this concept involves utilizing the fundamental matrix. Leveraging the fundamental matrix allows for the rectification of images and their alignment in a manner such that their epilines become congruent.

It is better if there are at least 8 pairings. This is when the RanSaC (Random Sample Consensus) approach comes in handy. RANSAC additionally takes into account the fact that not all matched features are credible [29]. It takes a random set of point correspondences, computes the fundamental matrix using them, and then evaluates its performance.

The algorithm chooses its best estimate when doing this for different random sets (typically 8-12). According to OpenCV's source code, the technique requires at least 15 feature pairs to function properly.

2.3.7. Epilines

As seen in the image below, the key points in the two images are located on the same epipolar lines.

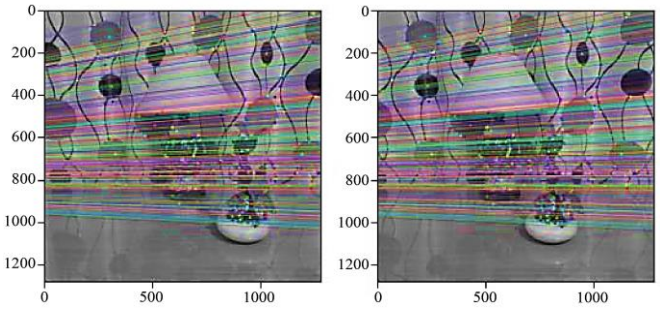

Fig. 14 Epilines in both images

2.4. Stereo Rectification

After completing all the preparatory steps, performing a stereo rectification using OpenCV is remarkably quick and simple.

For the rectification process, the matched points and the fundamental matrix are passed to the stereoRectifyUncalibrated() method, which is based on the work by Hartley et al. [29]. Subsequently, these transformations are applied to the images. The result is that both source images have now changed. The same two horizontal lines are drawn on top of the images. Features in both images are exactly on the same line in their respective image.

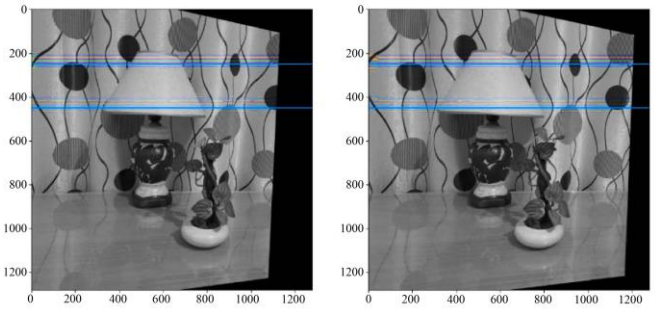

Fig. 15 Keypoints on the epipolar lines in rectified images that are the same in both images

Fig. 16 The disparity map generated based on the two images (right and left)

2.5. Stereo Matching

The final significant step in the process is stereo matching, and to achieve this, the Hirschmüller published Semi-Global Matching (SGM) [30] is applied. For this purpose, the SGM provided by the OpenCV library is employed. The algorithm's success crucially depends on configuring appropriate values for the scenes expected in OpenCV. The key parameters to be considered include block size, minimum and maximum disparity, and speckle.

2.6. Disparity Map

In stereo rectification, the left and right stereo images are transformed to align the corresponding spots along the same horizontal scanline. The distance B between the left and right camera centers is shown in Figure 16. The coordinates $q_1(x_l, y_l)$ and $q_r(x_r, y_r)$ denote the left and right image planes, respectively, onto which a point Q is projected. Z is a representation of the separation between Q and the camera.

One significant result of stereo rectification is the y-coordinate values of related spots in both images being the same, meaning $y = y_l = y_r$. This indicates that related locations will have the same y-coordinate when comparing the corrected left and right images side by side.

This kind of y-coordinate alignment makes it easier to compute the differences between related points. The horizontal shift in position between a point in the left image and its equivalent point in the right image is referred to as disparity. When matching locations have the same y-coordinate, it is easier to calculate disparity since you can compare pixel intensities along a single horizontal line. This improves the algorithm's efficiency and streamlines it.

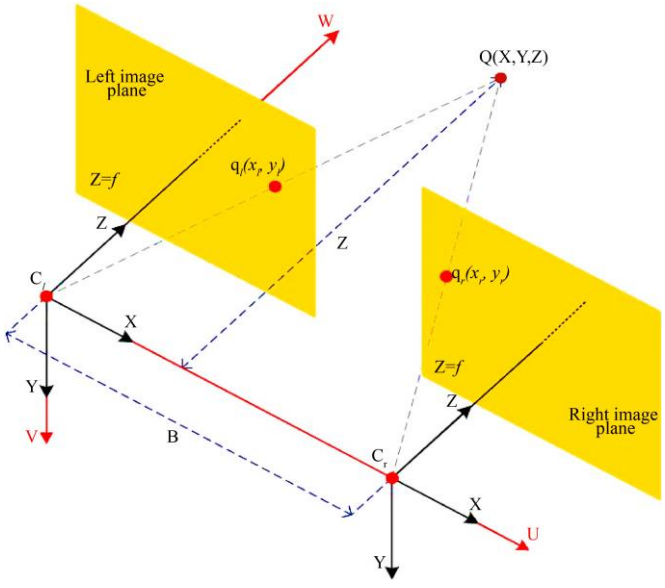

Fig. 17 Stereo triangulation scheme

The result of these computations is a disparity map, which shows the differences in depth between matching locations in the corrected left and right pictures. The disparity map is typically displayed as a 2D picture, with each pixel value representing the disparity or depth at that specific place. Among the many uses for this map are stereo vision system object identification, depth estimation, and 3D reconstruction.

Similar triangles allow us to establish the following equation:

$$\frac{B}{Z} = \frac{B - (x_l - x_r)}{Z - f}$$

The term "disparity" is often used to describe the difference between the left and right horizontal positions, denoted as x_l and x_r respectively. To make this equation more succinct, a new variable is introduced, which is called "d." This allows us to restate the equation in the following manner: $= x_l - x_r$. In this revised form, the equation is simplified and easier to work with as we replace the lengthier expression $x_l - x_r$ with the more concise "d." This not only streamlines the equation but also provides a clearer and more efficient representation of the relationship between these variables.

$$\frac{B}{Z} = \frac{B-d}{Z-f}$$

$$Z = \frac{f \times B}{d} \qquad (1)$$

The parameters 'f' and 'B' remain constant, representing fixed values in this context. However, the parameter 'd' holds particular significance as it represents the concept of "disparity." Disparity refers to the horizontal difference in location between a specific point in the left image and its corresponding position in the right image. By computing the value of 'd,' it becomes possible to ascertain the distance between an obstruction and the camera. In essence, this calculation of 'd' serves as a valuable tool for estimating the spatial relationship between objects in a stereo image pair, aiding in distance determination from the camera to objects of interest.

3. Discussion and Results

The scene contains two objects separated from each other by a distance of 30 cm. The closest object is positioned at a distance of 80 cm from the camera. In order to have an exploitable result and remove the imperfections on the resulting image, the disparity map post-filtering DisparityWLSFilter [32] is used to improve the results of the Stereo matching algorithm.

Equation (1) is used to get the distance, considering the observed disparity, between the obstacle and the camera. A 40 cm object was used in the studies and the distance at which it was positioned changed for every experiment.

Fig. 18 WLS post filter

More than sixty experiments were carried out with the object in front of the stereo vision system. The distances between these barriers ranged from 50 to 200 cm. The results of these experiments, which indicate the disparities between the measured distances and the real distances, are presented in Table 1. Additionally, Figure 8 visually represents the measured distances compared to the real distances of the obstacles.

The following formula is used to calculate the percentage error (PE) between the real distance (dr) and the measured distance (dm):

$$PE = \frac{|d_r - d_m|}{d_r}$$

In this formula, the absolute difference between the real and measured distances is divided by the real distance, and the result is multiplied by 100 to obtain the percentage error. This formula is commonly used to evaluate the accuracy of measurements by comparing them to the true values.

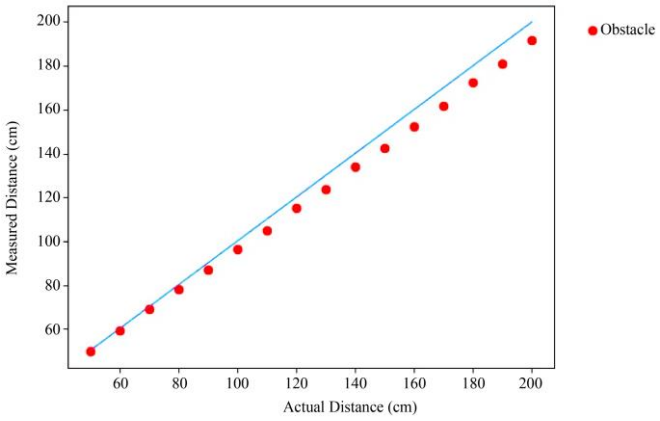

Fig. 19 The measured distances in relation to the real distances

Table 1. Distance measurement error

Real distance (cm)	Measured distance (cm)	Error (%)
50	49,68	0.65
60	59,27	1.225
70	69,17	1.187
80	78,14	2.325
90	87,16	3.158
100	96,34	3.658
110	104,95	4.592
120	115,22	3.987
130	123,57	4.948
140	134,05	4.248
150	142,46	5.028
160	152,50	4.689
170	161,83	4.807
180	172,30	4.278
190	181,07	4.698
200	191,49	4.255

The results that have been provided indicate that there is a significant correspondence between the distances measured during these studies and the actual distance values. The algorithm's precision in measuring distances is especially strong when an obstacle is placed in close proximity to the cameras. Moreover, the inaccuracy stays below 5% at distances up to 190 cm.

It is significant to remember that the observed disparity and the distance between the obstacle and the camera are inversely correlated; that is, as the distance reduces, the observed disparity increases. According to the study's experimental data, the average accuracy for obstacles located between 50 and 200 cm was 97.38% when comparing the measured and actual distances. For both, the lowest accuracy of 95.14% was noted.

Table 2. Comparing our algorithm with previous research

Real distance (cm)	Accuracy (%)	[33]	[34]	[35]
50	99.35	-	-	-
60	98.775	92.41	88.00	
70	98.813	96.75	92.00	98.57
80	97.675	99.41	98.00	-
90	96.842	99.98	98.00	-
100	96.342	99.04	98.00	-
110	95.408	99.65	99.00	-
120	96.013	99.52	98.50	95.83
130	95.051	99.79	99.00	-
140	95.752	99.49	99.00	-
150	94.972	-	98.00	-
160	95.311	-	97.50	-
170	95.193	-	97.50	-
180	95.722	-	97.00	-
190	95.301	-	95.50	-
200	95.745	-	95.00	-

It is necessary to complete several steps to obtain a depth map. Calculations are required at each stage. In order not to repeat all these calculations in the rectification part, the parameters obtained during the first rectification are saved to be able to be reused. This stereoscopic camera system created from two cameras of the same model is intended to create a real-time depth map; the operations described above will be repeated several times. The video taken from the two cameras will be transformed into image sequences, and this process will be repeated in real-time.

The support, which connects the two cameras, must not allow them to move to ensure any change in their position and orientation, which will affect the result. Another problem encountered is that the two cameras do not have a large field of view; this prevents a large number of key points from being obtained, resulting in a generally inaccurate result. Using another camera model with a wide field of view will guarantee a good result in most cases. The quality of the image taken also affects the result obtained. The model used in this system does not use a high-resolution camera. Hence, it reduces the chances of having a good result, which further complicates the operation of comparison between the two images taken by these two cameras. The result of matching key points is generally incorrect in some situations, for example, in the case where the image does not contain many key points.

4. Conclusion

This research paper presents the development of an obstacle detection system that employs stereoscopy as a critical tool to aid visually impaired individuals during their indoor navigation.

The proposed system aims to enhance the safety and autonomy of visually impaired individuals by providing real-time data and preventing collisions or accidents. By utilizing depth maps generated from stereo camera inputs, the system can accurately detect the proximity of obstacles and issue alerts, enabling users to adapt to diverse scenes for a safer experience and more independent navigation.

Since this system was created from scratch, it obviously needs improvements. Additional adjustments and tests can be accomplished in order to improve model performances.

In future work, the researchers aim to implement a different camera model to rectify errors induced by the current camera model. As well as using new features that can allow easy object detection and recognition in the user's environment, allowing them to not only detect objects but also recognize or classify them.

References

[1] Blindness and Vision Impairment, World Health Organization, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment

[2] Arjun Pardasani et al., "Smart Assistive Navigation Devices for Visually Impaired People," *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, pp. 725-729, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[3] B. Vamsi Krishna, and K. Aparna, "IoT-Based Indoor Navigation Wearable System for Blind People," *Artificial Intelligence and Evolutionary Computations in Engineering Systems*, pp. 413-421, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[4] Rahul Dhod et al., "Low Cost GPS and GSM Based Navigational Aid for Visually Impaired People," *Wireless Personal Communications*, vol. 92, pp. 1575-1589, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[5] G. Balakrishnan, and G. Sainarayanan, "Stereo Image Processing Procedure for Vision Rehabilitation," *Applied Artificial Intelligence*, vol. 22, no. 6, pp. 501-522, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[6] Larry Matthies et al., "Computer Vision on Mars," *International Journal of Computer Vision*, vol. 75, pp. 67-92, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[7] Faleh Alqahtani et al., "Detection and Tracking of Faces in 3D Using a Stereo Camera Arrangements," *International Journal of Machine Learning and Computing*, vol. 9, no. 1, pp. 35-43, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[8] Abdulkadir Akin, "*Real-Time High-Resolution Multiple-Camera Depth Map Estimation Hardware and its Applications*," Thesis, Infoscience EPFL Scientific Publications, pp. 1-221, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[9] Cécile Riou et al., "Calibration and Disparity Maps for a Depth Camera Based on a Four-Lens Device," *Journal of Electronic Imaging*, vol. 24, no. 6, pp. 1-11, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[10] Taih´u Pire et al., "Stereo Vision Obstacle Avoidance Using Depth and Elevation Maps," *Laboratorio de Robotica y Sistemas Embebidos*, pp. 1-3, 2012. [Google Scholar] [Publisher Link]

[11] Cristina Cattaneo, G. Mainetti, and R. Sala, "The Importance of Camera Calibration and Distortion Correction to Obtain Measurements with Video Surveillance Systems," *Journal of Physics: Conference Series*, vol. 658, pp. 1-8, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[12] V. Dwivedi, A. Kothari, and K. Raviya, "Depth and Disparity Extraction Structure for Multi View Images Video Frame-A Review," *European Journal of Academic Essays*, vol. 1, no. 10, pp. 29-35, 2014. [Google Scholar] [Publisher Link]

[13] Chao Yang, Fugen Zhou, and Xiangzhi Bai, "3D Reconstruction through Measure Based Image Selection," *2013 Ninth International Conference on Computational Intelligence and Security*, pp. 377-381, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[14] Ajay Kumar Mishra et al., "3D Surveillance System Using Multiple Cameras," *Videometrics IX*, vol. 6491, pp. 11-18, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[15] Vincent Nozick, "Camera Array Image Rectification and Calibration for Stereoscopic and Autostereoscopic Displays," *Annals of Telecommunications*, vol. 68, pp. 581-596, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[16] Andrey V. Kudryavtsev, Sounkalo Dembélé, and Nadine Piat, "Stereo-Image Rectification for Dense 3D Reconstruction in Scanning Electron Microscope," *2017 International Conference on Manipulation, Automation and Robotics at Small Scales (MARSS)*, pp. 1-6, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[17] Charles Loop, and Zhengyou Zhang, "Computing Rectifying Homographies for Stereo Vision," *1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, pp. 125-131, 1999. [CrossRef] [Google Scholar] [Publisher Link]

[18] Andrea Fusiel lo, "Tutorial on Rectification of Stereo Images," *Dipartimento di Matematica e Informatica Università di Udine*, pp. 1-12, 1999. [Google Scholar] [Publisher Link]

[19] Ruichao Xiao et al., "DSR: Direct Self-Rectification for Uncalibrated Dual-Lens Cameras," *2018 International Conference on 3D Vision (3DV)*, pp. 561-569, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[20] David Lowe, "Object Recognition from Local Scale-Invariant Features," *Proceedings of the Seventh IEEE International Conference on Computer Vision*, vol. 2, pp. 1150-157, 1999. [CrossRef] [Google Scholar] [Publisher Link]

[21] Vineetha Vijayan, and K.P. Pushpalatha, "FLANN based Matching with SIFT Descriptors for Drowsy Features Extraction," *2019 Fifth International Conference on Image Information Processing (ICIIP)*, pp. 600-605, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[22] David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91-110, 2004. [CrossRef] [Google Scholar] [Publisher Link]

[23] Tayeb Basta, "Is the Fundamental Matrix Really Independent of the Scene Structure?," *International Journal of Signal Processing, Image Processing and Pattern Recognition*, vol. 7, no. 5, pp. 149-168, 2014. [CrossRef] [Google Scholar] [Publisher Link]

[24] Justinas Miseikis et al., "Automatic Calibration of a Robot Manipulator and Multi 3D Camera System," *2016 IEEE/SICE International Symposium on System Integration (SII)*, pp. 735-741, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[25] Jiang Hong-zhi et al., "Phase-Based Stereo Matching Using Epipolar Line Rectification," *Guangxue Jingmi Gongcheng*, vol. 19, no. 10, pp. 2520-2525, 2011. [CrossRef] [Google Scholar]

[26] Fran ̧cois Darmon, and Pascal Monasse, "The Polar Epipolar Rectification," *Image Processing on Line*, vol. 11, pp. 56-75, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[27] Richard Hartley, and Rajiv Gupta, "Computing Matched-Epipolar Projections," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 549-555, 1993. [CrossRef] [Google Scholar] [Publisher Link]

[28] Rahul Raguram, Jan-Michael Frahm, and Marc Pollefeys, "A Comparative Analysis of RANSAC Techniques Leading to Adaptive Real-Time Random Sample Consensus," *European Conference on Computer Vision*, pp. 500-513, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[29] Richard I. Hartley, "Theory and Practice of Projective Rectification," *International Journal of Computer Vision*, vol. 35, pp. 115-127, 1999. [CrossRef] [Google Scholar] [Publisher Link]

[30] Monika Gupta, and Sapna Malik, "Dual Dynamic Programming with Quantization for Disparity Map," *SSRG International Journal of Computer Science and Engineering*, vol. 4, no. 3, pp. 5-9, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[31] Heiko Hirschmüller, "Stereo Processing by Semiglobal Matching and Mutual Information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 2, pp. 328-341, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[32] CV::ximgproc::DisparityWLSFilter Class Reference, Open Source Computer Vision, 2023. [Online]. Available: https://docs.opencv.org/3.4/d9/d51/classcv_1_1ximgproc_1_1DisparityWLSFilter.html

[33] Serdar Solak, and Emine Doğru Bolat, "A New Hybrid Stereovision-Based Distance-Estimation Approach for Mobile Robot Platforms," *Computers and Electrical Engineering*, vol. 67, pp. 672-689, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[34] Fredy Mart´ınez, Edwar Jacinto, and Fernando Mart´ınez, "Obstacle Detection for Autonomous Systems Using Stereoscopic Images and Bacterial Behaviour," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 2, pp. 2164-2172, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[35] Emre Dandil, and Kerim Kürşat Çevik, "Computer Vision based Distance Measurement System Using Stereo Camera View," *2019 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, pp. 1-4, 2019. [CrossRef] [Google Scholar] [Publisher Link]