Obstacle Detection Algorithm by Stereoscopic Image Processing

Navigation Assistance for the Blind and Visually Impaired

Adil Elachhab[✉] and Mohammed Mikou

Laboratory: Analysis of Systems and Information Processing,
Faculty of Sciences and Techniques, University Hassan I,
BP 577, 26000 Settat, Morocco
adil.elachhab@gmail.com, mohammed.mikou@uhp.ac.ma

Abstract. The continued growth in the number of blind or visually impaired people calls for an urgent need to invest all efforts to improve the quality of life for this population. This paper presents a work forming part of the technical study of new aid to the autonomous and secure movement for blind and partially sighted persons. These are intelligent glasses that capture environmental information, via a pair of stereoscopic cameras. In the case where an obstacle is present in the path of visual impairment, it will be detected by the stereoscopic sensor and an obstacle avoidance algorithm incorporated in a laptop computer. The final result, which provides key data to describe its environment, is transmitted to the user via headphones. Under this goal, we have developed a specialized processing algorithm that uses a stereoscopic vision approach and consists to determine the distance to obstacles and their size from two views of the same scene.

The results provided by the algorithm seem to be satisfactory in terms of computing time required and effective operating range.

Keywords: Obstacle detection · Visually impaired · Stereoscopic vision
Subpixel · Dynamic programming · Pyramidal image · Image segmentation
Super-pixels

1   Introduction

A visually impaired encounters a difficulty to reliably map the depth of the environment during its travels to an unknown destination. This limitation of accessibility, therefore, reduces its autonomy. A thorough review of aid systems based on RFID (Radio Frequency Identification) [1, 2], GPS (Global Positioning Systems) [3, 4], and ultrasound [5, 6], shows that despite efforts, these systems do not always provide a safe and independent mobility solution for the visually impaired. It appears necessary to find alternatives to determine the position of the obstacles continuously and in real time, and to extract the information necessary for navigation.

The arrival of stereoscopic vision system has been a source of great hope, because it is one of the most promising tracks in the navigation aid systems supposed to locate with certainty an obstacle and guide a blind or partially sighted person through his path [7, 8]. Stereoscopic vision systems extract depth information from two images acquired by two separate cameras placed at a distance B. The two cameras capture the scene that is then transmitted to a processing unit on which the proposed algorithm is integrated [9, 10]. The result of the processing device is further communicated to the user through an audible voice through earphones. In this context, we propose a new obstacle detection algorithm based on processing a pair of images taken by two synchronized cameras with parallel optical axes. In these conditions, two issues appear: how to successfully recover the objects depth in a scene in order to estimate the distance traveled before touching an obstacle? How can both get this reliably distance information, and within a reasonable time.

In the previous work [11–13], we proposed an algorithm that determines the obstacles position for both large and small obstacle-camera distances, evaluating the disparities between the two stereoscopic images by the algorithm "Block Matching" using the *SAD* criterion (Sum of Absolute Differences). Nevertheless, the results achieved are encouraging in terms of accuracy and reduced computational load, but they are not suitable for real-time applications. In this study, we propose to improve our algorithm by first considering the separation by segmentation the obstacles of the background before image treatment and the integration of advanced optimization approaches, resulting in a fast stereoscopic image processing. We also propose using this algorithm to determine the limit positions for the obstacle location, for which the visually impaired will be able to retrieve information about the obstacle in an efficient manner, allowing them to avoid it. The obstacle size in pixels^2 in the image is determined by the "Bounding Box" property of the predefined function "Regionprops" of MATLAB that locates and delimits obstacles with an encompassing box to retrieve their dimensions. The calibration curve established, giving the proportionality between the area in pixels^2 of the obstacle in the image and its real surface area in m^2, made it possible to evaluate the actual dimensions.

2   Materials and Methods

The stereoscopic image pairs are captured by means of an acquisition device that consists essentially of two digital cameras mounted on a linear and rigid support, which allows to properly stabilize and quickly adjust the distance between both objectives through a horizontal rail. The algorithm provides information on the depth and size of the obstacle in a reasonable time, enabling the blind or visually impaired to make the necessary decisions for his movement before approaching the obstacle. The proposed algorithm has been coded and tested in a MATLAB environment, Firstly, on nearby obstacles and small dimensions (Cameras-obstacle distance ranging from 1 m to 4 m with a step of 5×10^{-1} m); Front surface area varies between $4,200 \times 10^{-4}$ m^2 and $4,148 \times 10^{-2}$ m^2. Thereafter, the test is carried out on distant obstacles; Cameras-obstacle distance ranging from 5 m to 30 m with a step of 5 m. In this case, the area of the treated obstacles varies from $1,1088$ m^2 to $1,836 \times 10^1$ m^2.

2.1   Automatic Obstacle Detection

The algorithm used introduces a new obstacle extraction technique that requires three steps: The input image is first segmented into homogeneous color elements by "super pixel" exploiting the SLIC method. Next, we compute the contrast levels that evaluate the uniqueness and spatial distribution of these elements. Finally, we derive a binary mask that distinctly separates the obstacle of interest from the image background.

2.1.1   Image Segmentation Using SLIC

The pixels provide extremely local information that can not be used without intensive processing. By contrast, superpixels produce less localized information that contributes to the partition of an image into significant regions. The SLIC method adapts a k-means clustering approach [14] to efficiently generate compact and uniform superpixels that have a similar color distribution, subsequently forming an over-segmentation of an image. It performs a grouping of pixels in $5D$ space defined by the values L, a, b of the color space "L^*, a^*, b^*", which comprises all perceptible colors by the human eye and the pixel coordinates x, y. "L^*, a^*, b^*" space is used to capture the colors so that two colors perceived as similar are adjacent; very different colors are distinctly distant.

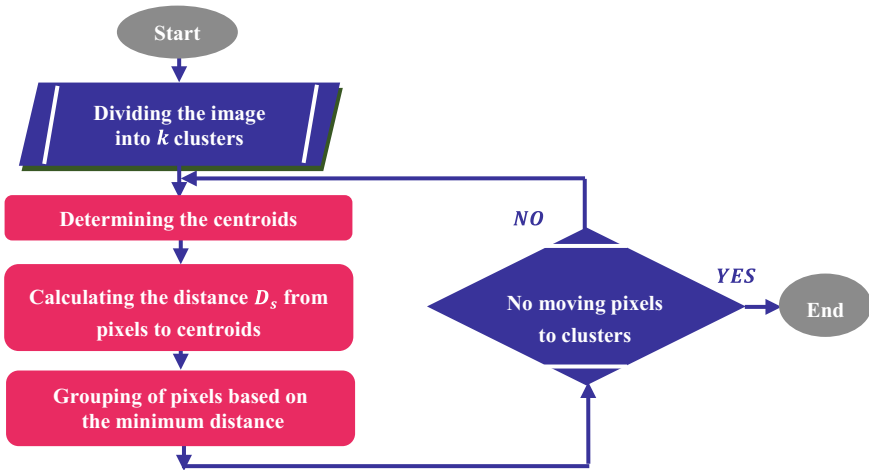In Fig. 1, we present the main steps of the SLIC algorithm implemented in *MATLAB*:

Fig. 1.   Schema of the *SLIC* algorithm using $k − means$ clustering

1. Dividing the image of n pixels into k groups sharing the same color, called clusters, with an approximate size of each cluster is n/k pixels. Thereafter, for superpixels of equal size, there is a cluster center at each grid interval $S = \sqrt{n/k}$.

2. Taking initial values for the cluster centroids $C_i = [L_i, a_i, b_i, x_i, y_i]^T$ where i varies from 1 to k.
3. Determining the distance of each pixel to the centroids using Euclidean distance D_s defined as follows [14]:

$$D_s = d_{lab} + \frac{m}{S} d_{xy}$$

With:

$$d_{lab} = \sqrt{(l_i - l_j)^2 + (a_i - a_j)^2 + (b_i - b_j)^2} \ \ and \ \ d_{xy} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

The variable m is introduced to control the compactness of a superpixel. The cluster is compact and spatial proximity is accentuated for a high value of m.

4. Grouping of pixels according to the minimum distance (search for the closest centroid).
5. Identifying new cluster centroids based on new accessions and repeating steps 3 and 4 until the pixels are no longer moved to another cluster. Thus, the computation of $k - means$ clustering reaches its stability and no iteration is required.

2.1.2 Local Contrast Difference

To extract the object of interest, the superpixels of the regions created by *SLIC* process are treated using the difference in average intensity between a pixel and its neighbors, each element is therefore represented by the average contrast of pixels belonging to it. Thus, depending on the uniqueness in the spatial distribution of superpixels, we produce a contrast map that uniformly covers the objects of interest and systematically separates the front and the background.

Objects of all sizes are present in the segmented image. To take the great foreground object constituting the obstacle, we label the connected regions, then we extract the surface of each. Through a thresholding according to these surfaces, we find the largest region by eliminating all the others. To determine the surface $(L \times l)$ in *pixels*^{2} that translates the white pixels number belonging to the object, the binary image is first scanned pixel by pixel in the four directions from the barycenter position of the object in search of the transition between its shape and the background, then the difference between the coordinates of transition pixels at horizontal and vertical level is calculated in order to count the pixels number in the width "L" as well as in the length "l".

2.2 Depth Calculation for an Obstacle

The depth calculation of a $3D$ point from stereoscopic images is almost immediate if these images are rectified, as can be seen in the figure below (Fig. 2):
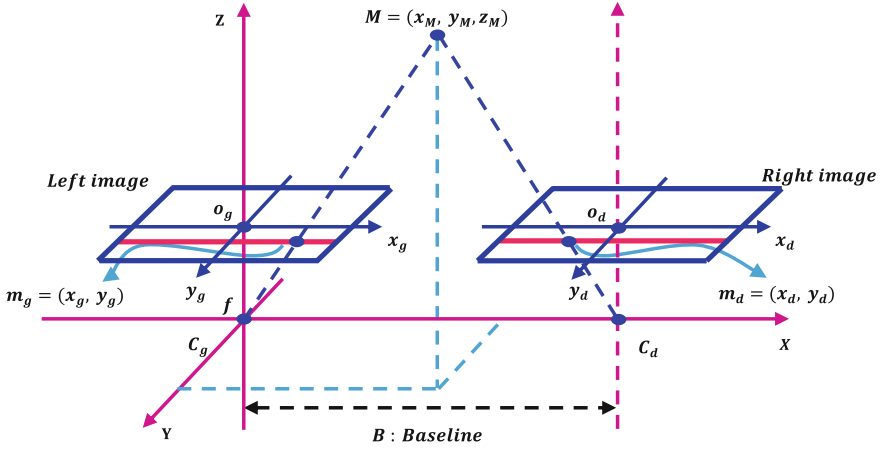
Fig. 2. Depth calculation of a 3D point for an aligned configuration of the cameras

A 3D point M has the projections:

$m_g = (x_g, y_g)$ in the frame $(o_g x_g y_g)$ associated with the left image,
$m_d = (x_d, y_d)$ in the frame $(o_d x_d y_d)$ associated with the right image.

The horizontal position x_g (resp. x_d) of the two points m and m', is given by:

$$x_g = f \frac{x_M}{z_M} \qquad \text{and} \qquad x_d = f \frac{x_M - B}{z_M}$$

So we have:
$$x_g - x_d = \left(f \frac{x_M}{z_M} \right) - \left(f \frac{x_M - B}{z_M} \right)$$

The difference between positions of the two projections in both images represents the disparity that can be found as shown in the following equation:

$$x_g - x_d = f \frac{B}{z_M}$$

The depth Z of $3D$ point is inversely proportional to the disparity measured on the image, defined as $d = x_g - x_d$.

$$z_M = \frac{B \times f}{x_g - x_d} = \frac{B \times f}{d} \qquad \text{With:} \quad x_g - x_d > 0$$

2.3   Stereoscopic Matching

Pixel matching into binocular stereovision consists to locate in two images of the same scene, taken at different positions, the pixel pairs which correspond to the projection of a same element.
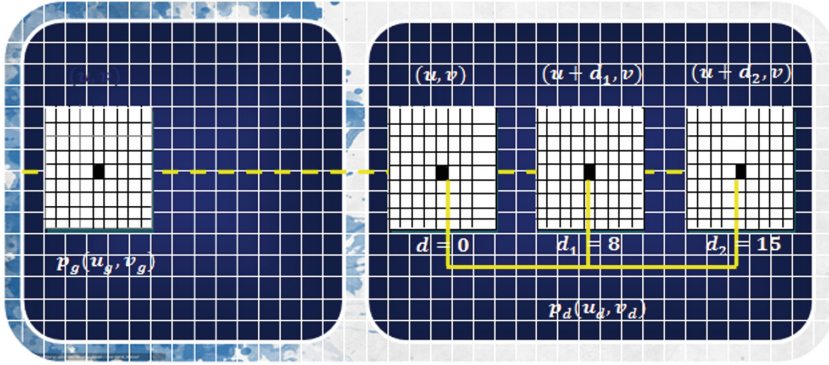
Fig. 3. Principle of the block matching algorithm

We can estimate the disparity by intensity correspondence of the two stereoscopic images. However, since there are often many intensity repetitions in one or more zones, pixel matching is unreliable. To solve this problem, the most commonly used method is to add neighborhood information to increase robustness. It is indeed the Block Matching algorithm, that we have employed in this work and which can be approached as an optimization problem, aimed at finding the best block within a research area. The corresponding costs for each pixel determine the probability of a correct match, the pixel selected is the one with the lowest cost (see Fig. 3).

A similarity measure is the degree of correlation between two windows, the correlation window centered on the pixel $p_g(u_g, v_g)$ to match from the left image and the sliding window centered on a potential pixel $p_d(u_d, v_d)$ located on the right epipolar of the other image I_d, of identical sizes $(2N+1) \times (2N+1)$. In this procedure, the disparity is obtained by minimizing the **SAD** criterion.

2.3.1 Disparity Map

To visualize the result of matching, we use an image called a disparity map. Each pixel of the map represents disparity amplitude: An important disparity will make the pixel appear lighter, and vice versa. Figure 4 shows an example of a stereoscopic image pair and the disparity map associated.

2.3.2 Refinement Process

In matching if the resolution is low, an error of a pixel generates an error of several meters in the distance calculation. Disparities values returned by the Block Matching are all integers, so that there is no smooth transition between regions of different disparities. In contrast the subpixel technique has been introduced to improve the disparity estimation quality for processed pixel. For example, if we have a 256×256 data matrix that we display at a resolution of 64×64 pixels, the data available to construct the image statistically exceed one sample per pixel. This means that instead of getting a location in the image in terms of pixel coordinates (x, y), which implies integer values for x and y, the location is calculated to eventually give positions of fractional pixels. Consequently, the positions of the correspondents are no longer integers causing a large amount of incorrect calculations, but real numbers [15].
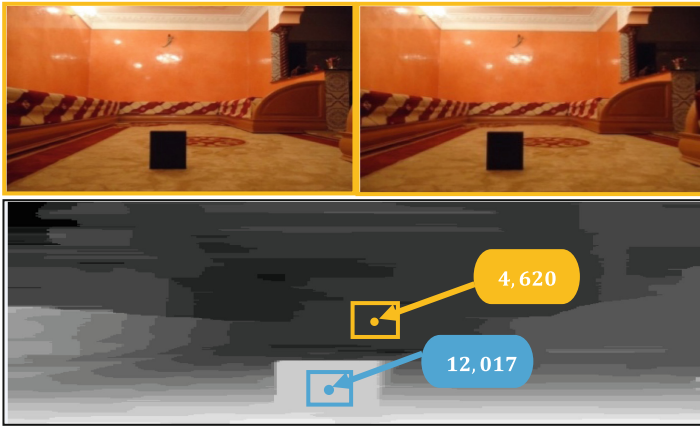
Fig. 4. Stereoscopic image pair with the associated disparity map

The problem of discerning optimal disparity estimates and eliminating false matches is solved by the dynamic programming algorithm which distinguishes the best match between two sets of pixels on homologous epipolar lines [16].

In view of the speed problems, the pairs of acquired stereoscopic images are sampled by means of a multi-resolution pyramid structure which is faster in the stereoscopic matching than the other without multi-resolution because the search range in each level of the pyramidal structure is considerably reduced [17].

2.4   Determination of the Obstacle Dimensions

The path of a blind or visually impaired person is always encumbered by obstacles of different forms which must be detected and identified in order to eliminate the risks of shocks. For this, we actively interested in estimating the actual dimensions of the surrounding obstacle whatever the distance, which is possible if we know its size in the image.

We use the *MATLAB* function Regionprops on the segmented image to locate obstacles with the Bounding Box property, defined as the smallest rectangle that can contain the dimensions or frame of an obstacle. The selection box around the obstacle with all abducted holes is drawn from its barycenter (gravity center) identified as follows: in a first step, we calculate the sum of 1 in the binary image which constitutes the total number of white pixels inside the obstacle form. In the next step, we return for all non-zero pixels the sum of its rows and columns coordinates in two variables, and consequently we determine the x and y barycenter coordinates while dividing the sum found by the integral number of 1.

Once the barycenter is identified, we take the column passing through this barycenter and we scan from its pixels each image line alternately to right and left, then as soon as the pixel value becomes 0, we store the coordinates of the last white pixel. Thus, we search in both directions for the pixel coordinates belonging to the obstacle contour giving a maximum distance with its gravity center. The same procedure is

repeated for the line that also passes through the barycenter, scanning in this case is done upwards and downwards. The coordinates of the four sides are used to draw the rectangle surrounding the obstacle in order to derive its width and height.

3   Results and Discussion

To facilitate the treatment in the Z-depth computation process of a $3D$ point, we must have rectified image pairs as inputs, which makes it possible to optimize and considerably simplify the pixels matching phase of the stereoscopic images. Epipolar rectification is a geometric transformation applied to the two stereoscopic images, and consists in switching to a sensor configuration that matching pixels lie on the same line (Fig. 5).
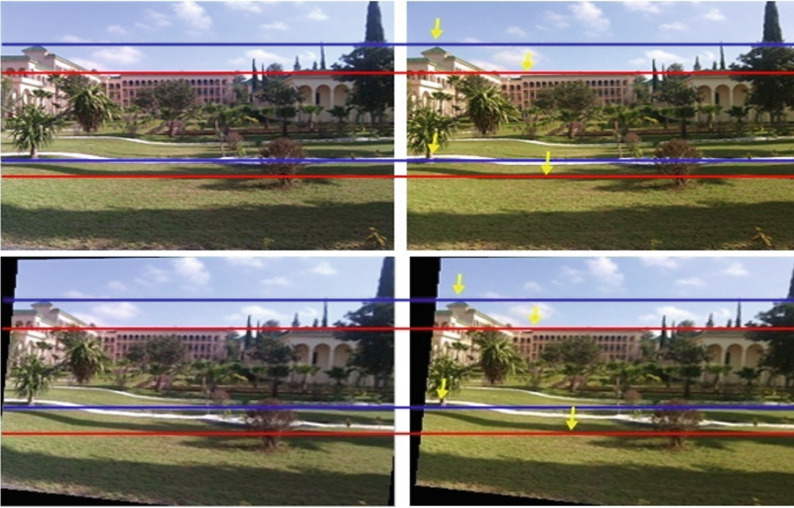

Fig. 5. Epipolar rectification: the first line shows images couple and its rectified version is on the second line

The red and blue lines in the above example show that the same point in the real world is on the same horizontal line in each rectified image. The rectification of stereoscopic images pair, therefore, enables to restrict the search space of the matches in a very important manner: we pass from an initial two-dimensional search space $2D$ to a 1D monodimensional space, because the disparity is only a difference of columns.

3.1   Evaluation of the Stereoscopic Algorithm

We tested in *MATLAB* a series of images for stereoscopic pair in order to determine the threshold distance below which the object distance between the images pair is apparent and easily detectable. The results obtained show that the shift observed in the

stereoscopic image is all the smaller as the objects are located very far from the sensor. Thus, the distinction between the images becomes imperceptible for 26 m. As a result, we can consider the distance 25 m as depth limit from which the objects have no visible disparity with the naked eye. While, the presented algorithm could discriminate the positional difference for each element of a scene up to a distance of 30 m.

3.2 Comment on the Main Results

Detection threshold corresponding to the smallest distance between the obstacle and the stereoscopic sensor for which the algorithm is able to determine the position and size of the obstacle is 0,5 m. When the distance falls below this value, both cameras do not cover exactly the same parts of the scene.

We can easily detect an obstacle located at a distance less than 0,5 m by bringing the optical axes of two cameras to converge on this obstacle, but in this case it is necessary to apply a transformation at both images so that their epipolar lines are parallel and aligned. In our algorithm, we avoided this costly solution in computing time by adopting a parallel configuration of cameras pair.

Detection limit giving the greatest distance from the obstacle to be detected is set at 30 m. Beyond this maximum detection value, we have not noticed a tangible difference between stereoscopic images pair, which means a disparity equal to 0 making it impossible to evaluate the camera-obstacle distance.

A valid comparison of the performance and efficiency of our algorithm with those of stereoscopic algorithms adopted by existing navigation systems [18–20] is difficult due to a lack of detailed information on their accuracy and processing time of the stereoscopic image pair, but we can estimate that the proposed method in this paper has the advantage over previous ones to determine in addition to the obstacle distance, the dimensions of the latter. It offers a large detection range, which provides time for the visually impaired to act. The treatment does not concern the entire image but only the region of interest constituting the obstacle, an approach that significantly helps reduce the execution time of the algorithm.

3.3 Detection Capacity and Robustness of the Presented Algorithm

Spontaneous walking speed of the visually impaired is approximately equal to 0,6 m/s. Furthermore, when walking the safety distance to be checked to avoid obstacles was defined as 2,5 m.

Our algorithm puts 2,89 s on a Windows 7 machine in 64 bits for an Intel Core *i*7 processor of 3, 40 GHz with 8192 MB of *RAM*. This means that during this time the visually impaired traverses a distance of 1, 734 m, at which is added the distance 0,6 m traveled during the reaction time (approximately 1 s), but which however remains less than 2,5 m. Accordingly, the developed algorithm might be able to guide mobility of visually impaired with adequate security without jerky behavior.

4   Conclusions and Perspectives

The performance of an obstacle detection system is inherent to the power of the stereoscopic vision algorithm. In this paper, we presented a new algorithm for the time-limited and optimal evaluation of the obstacle position and its size, based on Block Matching method. The capacity of the proposed method was verified by tests in a number of real stereoscopic sequences captured from a stereoscopic camera under different scenarios and with various changes in obstacle scales. Thus, we can estimate that our navigation aid algorithm for people who are blind or partially sighted has been developed, while taking into account the time constraint, whereby compliance is as important as result accuracy.

At the current stage in the advancement of our study, different perspectives can be offered. The first point to improve concerns the experiments on the algorithm execution time, and is considered as an unavoidable step in the battle of producing distance data between the stereoscopic camera and obstacles, in real time. Another interesting research track to explore is to test our solution in real conditions. Thus, we are actively considering converting information on potential obstacles obtained by image processing methods to sound signals or vibrations.

References

1. Doush, I.A., Alshatnawi, S., Al-Tamimi, A.-K., Alhasan, B., Hamasha, S.: ISAB: integrated indoor navigation system for the blind. Interact. Comput. **29**(2), 181–202 (2017)
2. Tsirmpas, C., Rompas, A., Fokou, O., Koutsouris, D.: An indoor navigation system for visually impaired and elderly people based on radio frequency identification (RFID). Inf. Sci. **320**(1), 288–305 (2015)
3. Dhod, R., Singh, G., Singh, G., Kaur, M.: Low cost GPS and GSM based navigational aid for visually impaired people. Wirel. Pers. Commun. **92**(4), 1575–1589 (2017)
4. Khenkar, S., Alsulaiman, H., Ismail, S., Fairaq, A., Jarraya, S.K., Ben-Abdallah, H.: ENVISION: assisted navigation of visually impaired smartphone users. Procedia Comput. Sci. **100**, 128–135 (2016)
5. Song, J., Song, W., Cheng, Y., Cao, X.: The design of a guide device with multi-function to aid travel for blind person. Int. J. Smart Home **10**(4), 77–86 (2016)
6. Bahadir, S.K., Koncar, V., Kalaoglu, F.: Smart shirt for obstacle avoidance for visually impaired persons. In: Smart Textiles and their Applications. A volume in Woodhead Publishing Series in Textiles, pp. 33–70 (2016)
7. Rizzo, J.-R., Pan, Y., Hudson, T., Wong, E.K., Fang, Y.: Sensor fusion for ecologically valid obstacle identification: building a comprehensive assistive technology platform for the visually impaired. In: 2017 7th International Conference on Modeling, Simulation, and Applied Optimization, ICMSAO 2017, 26 May 2017, Article Number 7934891 (2017)
8. Rodríguez, A., Javier Yebes, J., Alcantarilla, P.F., Bergasa, L.M., Almazán, J., Cela, A.: Assisting the visually impaired: obstacle detection and warning system by acoustic feedback. Sensors (Switzerland) **12**(12), 17476–17496 (2012)
9. Al-Mutib, K., Mattar, E., Alsulaiman, M.: Implementation of fuzzy decision based mobile robot navigation using stereo vision. Procedia Comput. Sci. **62**, 143–150 (2015). Proceedings of the 2015, International Conference on Soft Computing and Software Engineering (SCSE 2015)

10. Yang, Y., Gao, M., Zhang, J., Zha, Z., Wang, Z.: Depth map super-resolution using stereo-vision-assisted model. Neurocomputing, Part C **149**, 1396–1406 (2015)
11. Elachhab, A., Mikou, M.: Feasibility study of a navigation aid system for the visually impaired, determination of the depth and size of an obstacle by stereoscopic images processing. Int. J. Eng. Res. Technol. (IJERT) **3**(4), 180–185 (2014)
12. Elachhab, A., Mikou, M.: Obstacle detection algorithm based on stereoscopic images, a navigation aid system for the visually impaired. Int. J. Sci. Eng. Res. **6**(4), 872–878 (2015)
13. Elachhab, A., Mikou, M.: Obstacle detection based on stereoscopic vision and the local contrast difference, autonomous navigation assistance for visually impaired. J. Next Gener. Inf. Technol. (JNIT) **7**(3), 1–12 (2016)
14. Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S.: SLIC Superpixels, EPFL Technical, report 149300, June 2010
15. Arce, E., Marroquin, J.L.: High-precision stereo disparity estimation using HMMF models. Image Vis. Comput. **25**(5), 623–636 (2007)
16. Hu, T., Qi, B., Wu, T., Xu, X., He, H.: Stereo matching using weighted dynamic programming on a single-direction four-connected tree. Comput. Vis. Image Underst. **116**, 908–921 (2012)
17. Aribi, W., Kalfallah, A., Elkadri, N., Farhat, L., Siala, W., Daoud, J., Bouhelel, M.S.: Évaluation de Techniques Pyramidales de Fusion Multimodale (IRM/TEP) d'Images Cérébrales. In: 5th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications, Tunisia, 22–26 March 2009
18. Kammouna, S., Parseihianc, G., Gutierreza, O., Brilhaulta, A., Serpaa, A., Raynala, M., Oriolaa, B., Macea, M.J.-M., Auvrayc, M., Denisc, M., Thorpeb, S.J., Truilleta, P., Katzc, B. F.G., Jouffraisa, C.: Navigation and space perception assistance for the visually impaired: the NAVIG project. IRBM **33**(2), 182–189 (2012). Numéro spécial ANR TECSANTechnologie pour la santé et l'autonomie
19. Costa, P., Fernandes, H., Martins, P., Barroso, J., Hadjileontiadis, L.J.: Obstacle detection using stereo imaging to assist the navigation of visually impaired people. Procedia Comput. Sci. **14**, 83–93 (2012). Proceedings of the 4th International Conference on Software Development for Enhancing Accessibility and Fighting Info-exclusion (DSAI 2012)
20. Bujacz, M.: Sonified stereovision travel aid for local navigation, the project is financed by the National Centre of Research and Development of Poland in years 2010–2013 under the grant NR02-0083-10. http://www.naviton.pl/