ISSN 2229-5518

A Stereo Vision-based Guidance System for the Visually Impaired People

Ashraf Anwar[1] and Ibrahim El Rube[2]

Abstract —This paper proposes a stereo vision-based algorithm for objects' depth estimation suggested for guiding visually impaired people. The proposed algorithm is a first step towards a complete system designed for warning visually impaired persons to avoid in front obstacles at a reasonable distance (1m-2m). First, the left and right images are captured by a stereo vision camera. After preprocessing the captured images, objects are detected and segmented using morphological operations then a stereo image rectification is performed. The inliers points differences obtained from epipolar constraint are used to calculate the maximum and the minimum disparity. The suggested algorithm in this paper is tested using images captured for different objects with various distances from the stereo camera. Experimental results show that the proposed algorithm for depth estimation works quite robustly and faster than other methods in the literature.

Index Terms—Stereo image system, depth estimation, visually impaired, fundamental matrix, inliers points.

1 Introduction

Obstacle avoidance is one of the important problems that faces visually impaired people, especially when they are walking in places they are not familiar to. While most of them depend on trained dogs or white cane for walking, various devices do exist to assist these people for navigation without accidents. The development of hardware and/or software tools for assisting impaired people and provide accurate and robust information is now rapidly growing. Many researchers and companies are involved in this providing solutions to this problem. Several efforts in providing visually impaired people with necessary tools have been made [1]. The products that fall in this area are called Electronic Travel Aids, which can be a simple cane or more advanced computer based aids. Some of these ETAs use ultrasonic sensors to detect the obstacles within the path of the impaired people. In such devices, the visually impaired person usually get warned from obstacles by sound or vibrations signals [2], [3]. Usually these devices are expensive and suffer from operational problems due to multiple ultrasonic reflections and restricted mobility. Currently, many research efforts are being focused in improving and proposing navigation and travelling aid systems for visually impaired people who implements digital video cameras as an essential part in these systems. Furthemore, a stereo vision system with two or more cameras can be used to provide more accuracy to the navigation system by obtaining (depth) distances to obstacles in front of the user. In stereo vision, two images are captured from two different cameras at the same time but from slightly different viewpoints. Using a camera as vision sensor is attractive due its operational easiness, in expenses, riches of information it can provide, its al-nougy to the human eye, and low power consumption [4]. The epipolar geometry relates all images captured for the same scene with the corresponding points in these images constrained to lie on pairs of conjugate epipolar lines [5]. Stereo matching algorithms, usually, try to allocate matching points in two images that capture a scene.

Depth estimation algorithms can be broadly classified into two approaches; the first approach estimates the depth based on the disparity map computation. This approach requires both the extrinsic and the intrinsic parameters of the used camera. These parameters are obtained via camera calibration in the preprocessing stage and before performing stereo image matching [6-11]. The second approach images are rectified from a set of matched points from both images [12, 13]. One of the algorithms widely used in point detection and matching is the SURF (Speeded-Up Robust Features) proposed by Herbert Bay, et. al, [14]. SURF is a fast algorithm for the detection and description similarity transformed feature points. Local features in SURF are found from the image gradient distribution of the interest points obtained from scale-space approach. An open source SURF library is introduced by Evans, C. [15].

The fundamental matrix is estimated using point correspondences between two images. The difficulties in estimating the fundamental matrix lies in fact that there are often a fair portion of mismatches in a given set of point correspondences. It is therefore important that the method used for estimating the matching points and the method used to estimate the fundamental matrix should be robust.

There are several algorithms for robust estimation of the fundamental matrix [5]: RANSAC (RANdom Sample Consensus) technique [16] is a simple and successful method, it removes the effect of outliers by using random sampling as a search engine for inliers in the data set, the one that maximizes the cost function is chosen. A group of robust methods are developed based on RANSAC. LMeds (Least Median of squares) and LTS (Least Trimmed Squares) are similar to RANSAC, except for the way to determine the best solution.

In order to get accurate matching results that helpful for estimation of the fundamental matrix, an image pre-

1,2Department of Computer Engineering, Faculty of Computers and Information Technology,
Taif University, Saudia Arabia
[1]Email: ashraaf@tu.edu.sa
[2]Email: ielrube@yahoo.com

ISSN 2229-5518

processing stage is needed for smoothing the objects.

G. Balakrishnan, et al.,[4] proposed an image processing stage to identify objects based on locating their bounda-ries, they used Canny's method in addition to dilation and erosion operation, but this operations increase the computation time, which is critically considered in real time applications.

In this paper, the objects in front of the stereo camera are isolated by means of morphological operations in pre-processing stage without using Canny'method, resulting in time faster in pre-processing stage. Then, we use SURF matching algorithm for collecting interest points from each image because it gives better matching results. The fundamental matrix is obtained and depth is estimated by calculating the disparity from the inlier points in the left image and in the right image. We use the surveyor robot SVS [17], in Fig. 1 to capture the stereo images .

This paper is organized as follows: section 2 presents the stereo vision. Material and Methods are illustrated in section 3. Experimental results are presented in section 4. Finally concludes this paper.

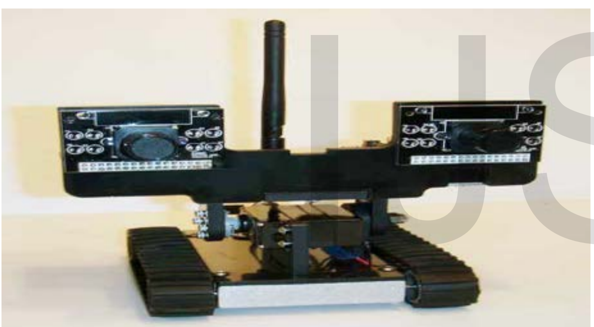


Fig. 1 SVS surveyor robot

2 STEREO VISION

2.1 Basics

The geometric basis key problem in stereo vision is to find corresponding points in stereo images. Corresponding points are the projections of a single 3D point in the different image

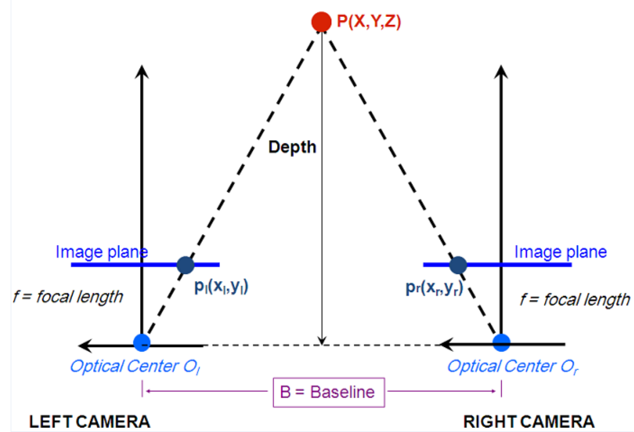


Fig. 2. Stereo vision basics

spaces. The difference in the position of the corresponding points in their respective images is called disparity, Fig. 2. Two cameras: Left and Right, Optical centers: Ol and Or. Virtual image plane is projection of actual image plane through optical centre. Baseline, B is the separation between the optical centers. Scene Point, P, imaged at Pl and Pr. Disparity, is the amount by which the two images of P are displaced relative to each other.

$$d = xr - xl. \quad (1)$$

$$\text{Disparity Depth,} \quad Z=Bf/d \quad (2)$$

In addition to providing the function that maps pair of corresponding images points onto scene points, a camera model can be used to constraint the search for corresponding image point to one dimension. Any point in the 3D world space together with the centers of projection of two cameras systems, defines an epipolar plane. The intersection of such a plane with an image plane is called an epipolar line. Every point of a given epipolar line must correspond to a single point on the corresponding epipolar line. The search for a match of a point in the first image therefore be reduced to a one-dimensional neighborhood in the second image plane.

3 MATERIALS AND METHODS

The main task in this paper is to get depth between the stereo camera and the obstacle in front of the user; this is the first step to design a stereo vision system for assisting visually impaired. In order to realize this task, we proceed according the following steps shown in Fig. 2

ISSN 2229-5518

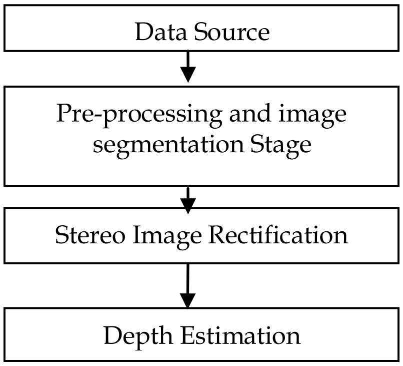


Fig. 3 Main proceeding steps

3.1 Data source.

We use surveyor robot SVS [18]. I t contains two SRV-1 Blackfin camera modules separated by 10.75 cm (4.25"). Each camera module includes:

- 500MHz Analog Devices Blackfin BF537 Processor (1000 integer MIPS), 32MB SDRAM, 4MB SPI Flash, JTAG, external 32-pin i/o header w/ 2 UARTS, 4 timers (PWM/PPM), SPI, I2C, 16 GPIO
- Omnivision OV9655 1.3 megapixel sensor with AA format header and interchangeable lens - M12 P0.5 format - 3.6mm f2.0 (90-deg FOV) or optional 2.2mm f2.5 (120-deg FOV)

3.2 Pre-processing and Object Segementation Stage

In this paper objects or obstacles are identified and are isolated based on its feature (area). The original image is shown in Fig. 4.

In order to locate the obstacles in the left image and right image, we propose the following image processing and morphological operations:

(1) Convert the Image to binary, Fig. 5
(2) Erosion and dilation operations are undertaken to smooth the objects, Fig. 6
(3) Measure the properties of image regions by applying “regionprops”, label the maximum area in descending order, locate the two maximums, and isolate the two maximums obstacles (left image), Fig. 7
(4) Re accumulate the two obstacles in one image, Fig. 8
(5) The resultant binary image with only objects are mapped with the corresponding gray scale image to derive new gray scale intensity images with only objects, Fig. 9
(6) These processes are performed for both left and right camera images to obtain two gray scale images with only objects

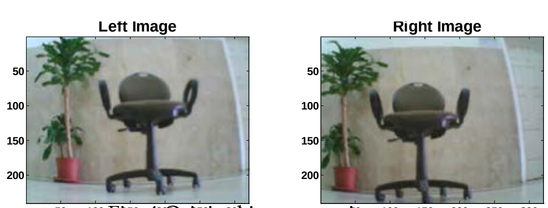


Fig. 4 Original image pair.

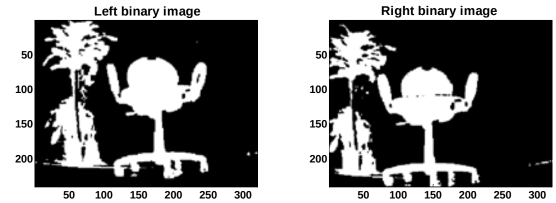


Fig. 5 Binary image pair.

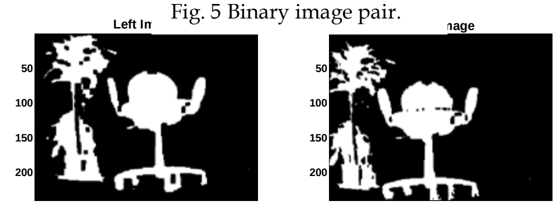


Fig. 6 Image pair after Erosion and dilation operations.

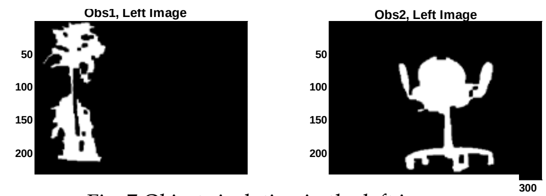


Fig. 7 Objects isolation in the left image.

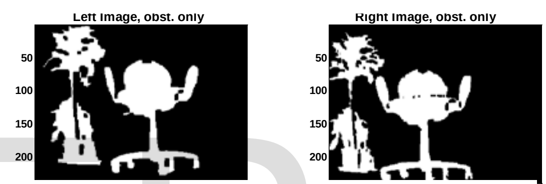


Fig. 8 Re accumulation process in the image pair.

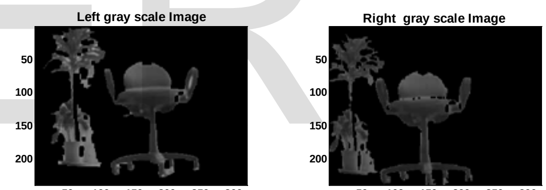


Fig. 9 Resultant gray scale image pair with objects only.

3.3 Stereo Image Rectification

The purpose of this stage is to image onto a common image plane in such a way that the corresponding points have the same row coordinates. It is often used as a pre-processing step for computing disparity. We proceed as the following flow chart illustrated in Fig.10

ISSN 2229-5518

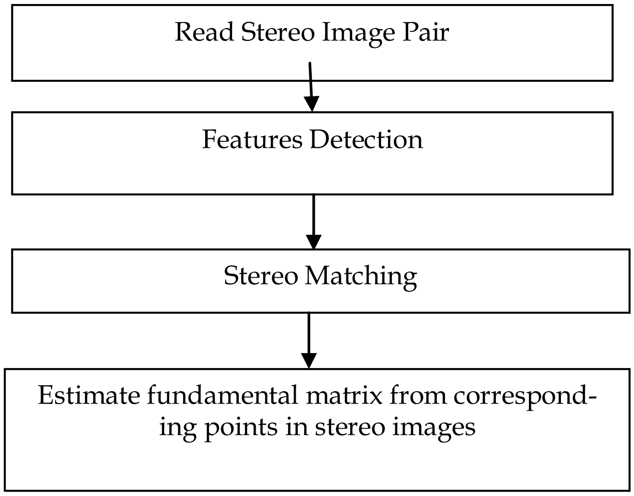


Fig.10 Flow chart of the stereo image rectification

(1) Features detection

Speeded Up Robust Features (SURF) is a robust local feature detector, first presented by Herbert Bay et. [14], that can be used in computer vision tasks like object recognition or 3D reconstruction. SURF is based on sums of 2D Haar wavelet responses and makes an efficient use of integral images.

The steps of features detection as follows:

- Interest points are selected at distinctive locations in the image, such as corners, blobs, and T-junctions. The most valuable property of an interest point detector is its repeatability, i.e. whether it reliably finds the same interest points under different viewing conditions.
- Next, the neighborhood of every interest point is represented by a feature vector. This descriptor has to be distinctive and, at the same time, robust to noise, detection errors, and geometric and photometric deformations.
- Finally, the descriptor vectors are matched between different images. The matching is often based on a distance between the vectors, e.g. the Mahalanobis or Euclidean distance. The dimension of the descriptor has a direct impact on the time this takes, and a lower number of dimensions is therefore desirable.

(2) Stereo matching.

The detection feature points must be matched. There exist several matching techniques based on various algorithms, e.g. Correlation (C), Normalized Cross Correlation (NCC), Sum of Squared Differences (SSD) and Sum of Absolute Differences (SAD) algorithms.

The SAD algorithm is one of the simplest of dissimilarity measures of the left and right stereo images corresponding with square windows. Hence, the algorithm was chosen for the proposed algorithm. It computes the intensity differences for each center pixel (i, j) in a window W(x, y) as follows:

$$SAD(x,y,d) = \sum_{(i,j)\in W(x,y)}^{N} |I_L(i,j) - I_R(i-d,j)| \qquad (3)$$

Where IL and IR are pixel intensity functions of the left and right image respectively, W(x, y) is square window that surrounds the position (x, y) of the pixel. The disparity SAD (x, y, d) calculation is repeated within the x-coordinate frame in the image row, defined by zero and maximum possible disparity dmax of the searched 3D scene. The minimum difference value over the frame indicates the best matching pixel, and position of the minimum defines the disparity of the actual pixel [10].

(3) The Fundamental matrix.

The fundamental matrix F is a 3×3 matrix which relates corresponding points in stereo images. In epipolar geometry, with homogeneous image coordinates, x and x′, of corresponding points in a stereo image pair, Fx describes a line (an epipolar line) on which the corresponding point x′ on the other image must lie. That means, for all pairs of corresponding points holds

$$X'FX = 0 \qquad (4)$$

The correctly matched points must satisfy epipolar constraints. This means that a point must lie on the epipolar line determined by its corresponding point. RANSAC: RANdom SAmple Consensus. It can be considered like a search engine. It selects repeatedly a random sample of 7 correspondences and computes the inliers realizing the fundamental matrix. The solution with the most inliers retained [5].

3.4 Depth Estimation

We can get disparity (max, min) from the inlier points in both the left image and in the right image, obtained as output from fundamental matrix

$$dis_{max} = max[\sum_{i=1}^{m} |I_{nL}(i,j) - I_{nR}(i,j)|] \quad (5)$$

$$dis_{min} = min[\sum_{i=1}^{m} |I_{nL}(i,j) - I_{nR}(i,j)|] \quad (6)$$

Where, I_{nL}, I_{nR} are the inlier points in the left and in the right image respectively

The estimated depth can be calculated as follows

$$R_{min} = \frac{F*B}{dis_{max}} \quad (7)$$

$$R_{max} = \frac{F*B}{dis_{min}} \quad (8)$$

Where, B base line between the left and right camera, focal Length

The min. depth accuracy is calculated according to the following formula:

$$Acc\ \% = \left|1 - \frac{Real\ min\ depth - estimated\ min.\ depth}{Real\ Depth}\right| 100 \quad (9)$$

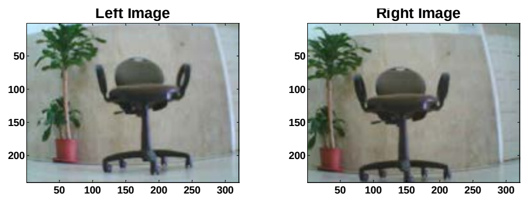
4 EXPERIMENTAL RESULT

4.1 Processing Steps.

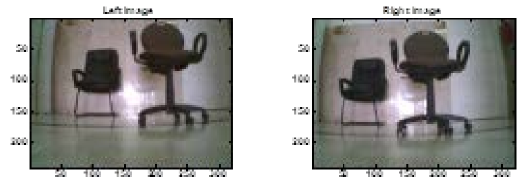
In order to assess the performance of the algorithm, we test on a set of images captured by stereo vision system of SVS surveyor robot. The tested image-pairs are shown in Fig. 11

The simulation is performed using matlab software (R2012a). The computer processor is Intel ® core TM, i5, M430, 2.27 GHz. The proposed algorithm is simulated and processed on every image pair A,B,C. We perform 50 run simulations for every image pair and then take the mean value. The experimental results for every image pair are tabulated in Table 1

We can deduce the inlier points in the left and in the right image by finding the inliers that meet the epipolar constraint.



Im-A



Im-B

Im-C

Fig. 11 Tested Image pairs

Table 1 Experimental results

Method	Im-A		Im-B		Im-C		Elapsed time (sec)
	R_{min}(cm)	R_{max}(cm)	R_{min}(cm)	R_{max}(cm)	R_{min}(cm	R_{max}(cm)	
	110	210	120	240	100	200	
Proposed pre-processing	113.7	194.4	116.6	229.4	101.9	202.2	1.2
Acc %	96.7	92.5	97.1	95.6	98.1	98.9	
Canny Method [ ]	115.6	189.7	122.8	193.9	105.6	193.3	1.56
Acc %	94.9	90.3	97.6	80.7	94.4	96.7	

4 Conclusions

An efficient and robust algorithm for depth estimation is presented in this paper. The algorithm estimates the objects' depth in stereo images using morphological operations and region properties. Stereo images are matched and rectified after locating correspondences between images by SURF and SAD distance. A comparison to a similar algorithm in the literature, but with Canny edge detection, shows that our algorithm is faster and more accurate in estimating objects' depth. The proposed algorithm is considered as a first step towards designing a stereo vision system for visually impaired people.

References

[1] F. Wong, R. Nagarajan, S. Yaacob, A. Chekima and N. E. Belkhamza, "Electronic travel aids for visually impaired – A guided tour", Proceedings of Conference in Engineering in Sarawak, Malaysia, pp. 377-382, 2000.

[2] C. Capelle, C. Trullemans, "A Real-Time Experimental Prototype for Enhancement of Vision rehabilitation Using Auditory Substitution", IEEE Trans. on Biomedical Engineering, Vol. 45, pp. 1279-1293, 1998.

[3] J. Borenstein and I. Ulrich, "The GuideCane – A computerized travel aid for the active guidance of blind pedestrians", Proceedings of IEEE International Conference on Robotics and Automation, Albuquerque, NM, pp. 1283-1288, 1997.

[4] G. Balakrishnan, et al., "A stereo image processing system for visually impaired.", Inernational journal of Information and communication Engineering 2:3, 2006

[5] Hartely, R.I., Zisserman, A., "Multiple View Geometry in Computer Vision. 2edn. Cambridge Press 2004

[6] Harkanwal, et al. (2011)," A robust area based disparity estimation technique for stereo vision applications", proceeding of the 2011 International conference on image processing.

[7] Meng Chen, et al.,"A method for mobile robot obstacle avoidance based on stereo vision", proceeding 10th IEEE International conference on components, circuits, devices, and systems, 2012

[8] Zhao Yong-guo, et al. "The obstacle avoidance and navigation based on stereo vision for mobile robot" proceeding international conference on optoelectronics and image processing., 2010

[9] Ming Bai, et al., "Stereovision based obstacle detection approach for mobile robot navigation", proceeding international conference on intelligent control and information processing, 2010.

[10] Patrik Kamencay, et al., "Improved depth map estimation from stereo images based on hybrid method", Radio engineering, vol. 21, NO. 1, pp. 70-78, 2012.

[11] Kanade, T., and M. Okutomi. "A stereo matching algorithm with and adaptive window' Theory and experiment. IEEE Transactions on Pattern Analysis and Machine Intelligence, (TPAMI) 16: 920-932, 1994.

[12] Wu, H.H., Yu, Y.H., "Projective rectification with reduced geometric distortion for stereo vision and stereoscopic video", Journal of Intelligent and Robotics Systems 42 , 71-94, 2005.

[13] Isgro, F., Trucco, E., "Projective rectification without epipolar geometry", Computer Vision and Pattern Recognition. 1, 119-128, 1999.

[14] Herbert Bay, H., A. Ess, T. Tuytelaars, and L. Van Gool., "Speeded-up robust features (SURF)". Computer Vision ECCV 2006, Vol. 3951. Lecture Notes in Computer Science. p. 404-417, 2006.

[15] Evans, C., "Notes on the Open SURF library". UK Publication Papers. Issue 1,Citeseer. p. 25, 2008.

[16] P.H.S Torr, D.W. Murray, "The development and comparisons of robust methods foe estimating the fundamental matrix", International journal of computer vision, 24 (3), 271-300, 1997

[17]http://www.surveyor.com/stereo/stere o_info.html

Dr. Ashraf Anwar received the PH.D degree in Electronics and Electrical Communication from Faculty of Engineering, Cairo University, Egypt in 2005. Currently, he is an Assistant Professor in college of computers and Information Technology, Taif University, Saudi Arabia. His research interests are mainly in the area of, computer vision, intelligent tracking, control, and robotics.