A Robust and Efficient Homography Based Approach for Ground Plane Detection

Ajay Mittal[1] and **Sanjeev Sofat**[2]

Submitted in December 2011; Accepted in February 2012

Abstract – This paper presents a homography based ground plane detection method. The method is developed as a part of stereo vision based obstacle detection technique for the visually impaired people. The method assumes the presence of a texture dominant ground plane in the lower portion of the scene, which is not severe restriction in a real world. SIFT algorithm is used to extract features in the stereo images. The extracted SIFT features are robustly matched by model fitting using RANSAC. A sample of putative matches lying in the lower portion of the image is selected. A fitness function is developed to select matches from this sample, which are used to estimate ground plane homography hypothesis. The ground plane homography hypothesis is used to classify the SIFT features as either belonging to ground plane or not. Image segmentation using mean shift and normalized cut is further used to filter the outliers and augment the ground plane. Experimental tests have been conducted to test the performance of the proposed approach. The tests indicate that the proposed approach has good classification rate and have operating distance range from 3 feet to 12 feet.

Index Terms - Ground Plane; SIFT; Electronic Travel Aid, Homography

1. INTRODUCTION

To get the perception of the environment around them, humans depends upon five senses- vision, hearing, smell, touch and taste. Among these, vision is undoubtedly the most dependable one. Most people cannot imagine what life would be, if they lose it. This is however, a hardcore reality for 45 million people worldwide, who are blind. World Health Organization (WHO) in the year 2010 has estimated that the worldwide count of Visually Impaired (VI) people is about 314 million and 45 million amongst them are completely blind [1].

VI people experience serious difficulties in leading an independent life, due to reduced perception of the environment [2]. The most obvious problem faced by VI people is in navigating the unknown environments without bumping into unexpected obstacles. Thus, obstacle detection is one of the major problems that need to be solved to ensure safe navigation for VI people.

The problem of obstacle detection may often be reduced to the problem of ground plane detection. With the ground plane detected, the other objects can be viewed as obstacles, if they protrude outside of the ground plane. In this paper, we have proposed a homography based ground plane detection method. The developed method is a part of stereo vision based obstacle detection technique developed for VI people. The proposed method assumes the presence of a texture dominant ground plane in the lower portion of the scene, which is not severe restriction in real world. The SIFT matches lying in the lower portion of the image and selected by a fitness function are used to generate a ground plane homography hypothesis. The generated homography hypothesis is used to classify matched SIFT features as either belonging to ground plane or not.

The paper is organized as follows: Section 2 presents an overview of related work. Section 3 explicates the theoretical background of homography. Section 4 presents the proposed approach in detail. The experimental results are given in Section 5. Section 6 makes the concluding remarks.

2. RELATED WORK

In context of navigation for a visually impaired user, obstacle can be defined as "anything that stops the progression of the user and/or requires the modification of his/her posture." For the past many years, the VI people have been relying greatly on the use of white cane during navigation. However, white cane has an inherent disadvantage. It cannot be used to obtain the information about the obstacles beyond its reach and hence cannot help the user in broad route planning.

Since 1960's, extensive research has been carried out for developing electronic devices, known as Electronic Travel Aids (ETAs), to assist VI people in autonomous navigation [3]. A number of ETAs that make use of radar, lidar and sonar technology [4-11] have been developed. However, their major disadvantages include: interference with the environment, difficult interpretation of the output signals, high power consumption, high acquisition price, poor angular resolution and incapability to detect small obstacles.

Vision-based ETAs have seen tremendous development in the recent years, largely due to the availability of low-cost cameras and compact yet high performance processors that support image processing. Being passive in nature, vision based aids have low power consumption and do not interfere with the environment. These systems work in the direction of capturing the image of an environment and mapping the image into sound or vibratory pulses without undertaking any image processing efforts to provide the information of objects in the scene. In general, background fills more area in an image frame than the objects and hence, conversion from unprocessed images will lead to information overload, with background details masking primary mobility information.

Automatic pre-processing to provide mobility data at a high level of abstraction, by eliminating detailed clutter but retaining essential mobility information, can alleviate the problem of information overload. Ground plane perception is the vital information for human mobility [12]. Gibson in [13] suggested

[1, 2] Department of Computer Science and Engineering, PEC University of Technology, Chandigarh, INDIA
E-mail: [1]ajaymittal@pec.ac.in and [2]sanjeevsofat@pec.ac.in

that "the spatial character of the visual world is given not by the objects in it, but by the ground and the horizon." Molton [14] developed a stereo-based mobility aid for partially sighted people by estimating ground plane using disparity information. Many approaches [15-25] for ground plane estimation for mobile robot and autonomous guided vehicle (AGV) navigation have been investigated by various researchers. These approaches rely mostly on the processing of different features attached to the ground planes: color [18, 19, 20], texture (lane markings) [18], disparity [14, 19, 20], v-disparity [21, 22], motion (optical flow [19, 23]), homography estimation [24, 25]

3. THEORETICAL BACKGROUND

There exist projective relationships between two viewpoints of a scene taken from a stereo rig. The corresponding points in stereo images, taken from uncalibrated cameras, are related by a fundamental matrix. If x and x' are the homogenous image coordinates of the corresponding points $\{x \leftrightarrow x'\}$ in a stereo image pair, **Fx** describes an epipolar line on which the corresponding point **x'** on the other image must lie. For each pair of corresponding points, the epipolar constraint is expressed as:

$$x'^{T}Fx = 0, \quad (1)$$

where F is a fundamental matrix. It is a3 by 3 matrix of rank 2 with seven degrees of freedom, hence it can be recovered from 7 point correspondences.

If a set of points lie in a plane, and they are imaged from two viewpoints, then the homogenous coordinates of the corresponding points $\{x_i \leftrightarrow xi'\}$ in the two images are related by a plane-to-plane projectivity or homography such that:

$$\lambda x_i = H x_i', \quad (2)$$

where **H** is a 3 by 3 matrix representing homography and λ is a scalar. Since equation 2 is valid up to a scale factor, **H** has only eight degrees of freedom and it is normal practice to choose λ such that the element h_{33} in **H** is set to unity. To determine **H**, four corresponding non-degenerated coplanar points are required, since each point correspondence provides two independent constraints, thereby making H determination possible by standard linear methods. However in reality, with the data being non-perfect, more number of point correspondences should be used for the accurate estimation of H.

4. PROPOSED APPROACH

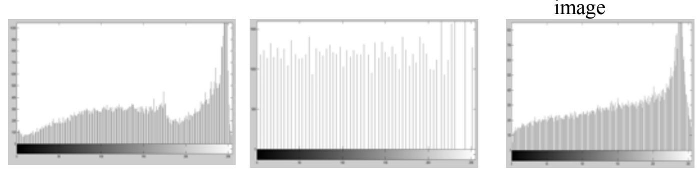
The steps involved in the proposed approach are described in the following sub-sections in detail:

4.1 Image Processing

The stereo images grabbed with low cost web cameras, arranged to form a stereo rig, often contain noise. The image noise can be eliminated by using Gauss filter. Contrast- Limited Adaptive Histogram Equalization (CLAHE) algorithm [26] is then applied to enhance the partial contrast and selectively highlight the obvious features, so that the resultant images are more conducive to feature extraction. As opposed to histogram equalization, CLAHE operates on small regions in the image. It enhances the contrast of each region and eliminates the artificially induced boundaries in the neighboring regions by using bilinear interpolation.

Figure 1 shows the result of image preprocessing stage.

(a) Original image (b) Histogram equalized image (c) Adaptive Histogram Equalized image



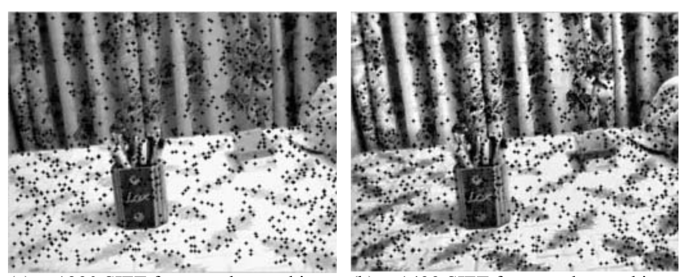
(d) Histogram of (a) (e) Histogram of (b) (f) Histogram of (c)

Figure 1: "The results of preprocessing stage"

4.2 Feature Extraction

The preprocessed image is given to SIFT feature extractor. Lowe proposed SIFT algorithm [27], which consists of four major stages: (1) scale-space extrema detection, (2) keypoint localization, (3) orientation assignment, and (4) keypoint descriptor.

Among various image features like corners, edge features, moment invariants, etc., SIFT features have been an obvious choice for the proposed approach because they are invariant to scale, orientation, and affine distortion, and partially invariant to illumination changes. Using SIFT algorithm, the keypoints extracted for Figure 1 (a) and (c) are shown in Figure 2.



(a) 1290 SIFT features detected in Original image (b) 1439 SIFT features detected in CLAHE applied image

Figure 2: "SIFT features extracted for Figure 1(a) and (c)"

The more number of SIFT features in Figure 2(b) illustrates the advantage of the performed preprocessing step.

4.3 Feature Matching

The SIFT features extracted in the left and right stereo images are matched using the procedure prescribed in [27, 28]. The matches that are too ambiguous are rejected. Matching follows a nearest neighbor approach. The feature F_i in the left image matches feature F_j in the right image, if the distance

$$d(F_i,F_j) < \tau d(F_i,F_k), \text{ for all } k \in I_r;\ j \neq k \quad (3)$$

where τ is a threshold. We have selected the value of the threshold τ to be 0.5. The initially obtained matches are used to estimate homography and fundamental matrices between pair of images using RANSAC [29]. The set of matches which fits a certain model (homography or fundamental matrix) are considered as inliers for that model. In order to further increase the robustness, outlier rejection rule, called X84 [30], is applied. The robustified inliers are used to re-estimate the parameters of the models. The best-fit model (homography or fundamental matrix) is selected according to the Geometric Robust Information Criterion (GRIC) [31]. The final matches are the inliers from the best-fit model.

4.4 Ground Plane Homography Hypothesis

Since, the algorithm assumes that the textured dominant ground plane lies in the lower part of the scene; a sample of putative matching points lying in the lower part of the image is selected. In our experiment, we have created the sample with the putative matches that lie in the lower 10% of the image. A heuristically designed predefined ground plane mask, as in [20], or a trapezoidal region in the lower central part of the image, as in [32], can also be used to select a sample of putative matches that lies on the ground plane.

The selection of four initial points from this sample to estimate homography is of vital importance. These four initial points will influence the likelihood that they determine a valid homography. The selection of the four initial points from the sample is based on the following criteria [33]:

1. Selected points should not be too distinct.
2. Selected points should not be too close.
3. No three selected points should be collinear or near collinear.
4. Selected corresponding points should have a large disparity in position.

Practically more than four points are required for the accurate estimation of homography matrix H. Algorithm SelectBestN listed in Table 1, is used to select best N-points based on the fitness score, computed for each point according to the mentioned fitness criteria.

1 **Algorithm** SelectBestN(*Ml*, *M'l*, *n*, *Tc*, *Td, N*)
2 r←1
3 Form pairs of points (p_i,p_j),such that p_i, p_j€*Ml*, $1 \leq i \leq n$ AND i<j. Select a pair (p_i,p_j) and for every pair repeat step 4 to 8
4 Find midpoint *m* of the point p_i and p_j, fit a line *l* passing through these points and fit a line *q* perpendicular to line *l* and passing through the midpoint *m*
5 Find distances d(p_k,l), d(p_k, q) and d(p_k,m); for all p_k€*Ml* and k≠i,j
6 If distance $d(p_k,m)>Tc$ and $d(p_k,m)<Td$, set $i(k)=1$ else set $i(k) = -1$
7 Compute fitness score for each point *pk*: $fs(r,p_k)=0.3\times d(p_k,l)+0.3\times d(p_k,q)+0.2\times\delta(p_k)+i(k)\times 0.\ m\ 2\times d(p_k,)$
8 r←r+1
9 Select r, such that *median(fs(r,:))* is maximum.
10 Sort *fs(r,:)*. Return p_i and p_j for the selected value of r and topmost N-2 points from *fs(r)*

"ALGORITHM to select best N-points based on fitness scores"

Figure 3 shows the result of execution of SelectBestN algorithm. In the figure, the '+' markers are SIFT matches, 'o' markers is the sample lying in the lower 10% of the image. The '⊕' markers are selected by the SelectBestN algorithm for N=6.

Figure 3: "Result of SelectBestN algorithm"

4.4.1 Goodness of Homography Estimate

The determinant of the homography matrix signifies the goodness of a homography estimate [34]. If the determinant tends towards to zero, it suggests the arrival of degeneracy in the selected points. Table 2 lists the values of determinant of homography matrix, average distance error and median of distance errors for three different test executions.

Test Run No. 1				
No. of pts. used for H estimation	**4**	**6**	**8**	**All in the sample**
det(H)	1.04012	0.9950	0.8681	1.2627
Avg(dist(x-Hx'))	0.8857	0.8489	0.8237	1.1525
Median(dist(x-Hx'))	0.9280	0.9406	0.8691	1.1329
Test Run No. 2				
No. of pts. used for H estimation	**4**	**6**	**8**	**All in the sample**
det(H)	1.5900	1.4764	1.0605	1.0569
Avg(dist(x-Hx'))	0.9915	1.3620	0.6228	1.0633
Median(dist(x-Hx'))	0.8415	1.1692	0.5706	1.0265
Test Run No. 3				
No. of pts. used for H estimation	**4**	**6**	**8**	**All in the sample**
det(H)	**0.2389**	1.4246	1.1571	1.0793
Avg(dist(x-Hx'))	**7.5435**	1.4802	0.9292	1.1583
Median(dist(x-Hx'))	**8.4288**	1.3374	0.7643	0.9412

Table 2: "Goodness of Homography estimate for different test runs"

The results indicates that avg(dist(x-Hx')) is very high, if det(H) tends to zero. The homography estimate with minimum average distance error is selected for classifying SIFT features as either "belong to ground" or "not".

4.5 Classification of Sift Features

The corresponding points that lie on a plane shares the same homography, which is different from the homography for another plane. In reality the homography equation 3 is satisfied only approximately. A pair of corresponding points (x,x') is considered to agree with the ground plane homography hypothesis, H, if for some threshold ε,

$$dist(x,Hx') < \varepsilon \qquad (4)$$

Equation 4 is used to classify the matched SIFT features as belonging to ground plane or not. Figure 4 shows the matched SIFT features that have been classified as belonging to ground plane, with the value of ε being kept 5.

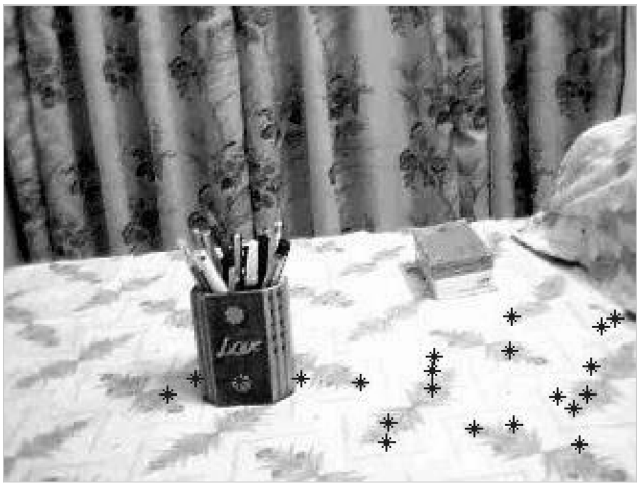


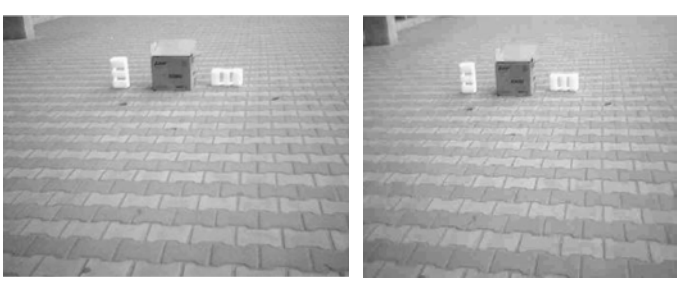
Figure 4: "SIFT features belonging to ground plane"

4.6 Augmenting Ground Plane

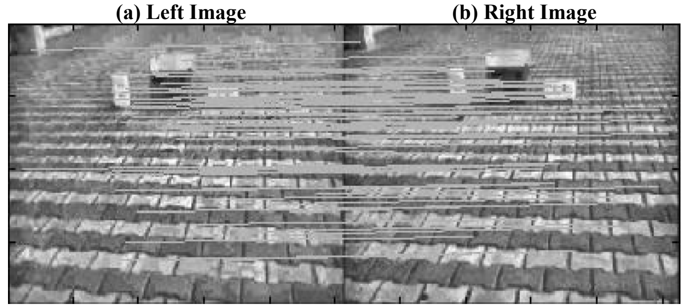
It is likely that that the feature points classified as lying on the ground plane are in the small region of the whole plane. In such circumstances, the region enclosing the classified feature points would not be accepted. We have addressed this problem by using image segmentation algorithm based on mean shift and normalized cuts [35]. The segmented region which encloses maximum number of feature points classified as lying on ground plane is finally labeled as ground plane. This step has an additional advantage of filtering outliers.

5. EXPERIMENTAL TESTS AND RESULTS

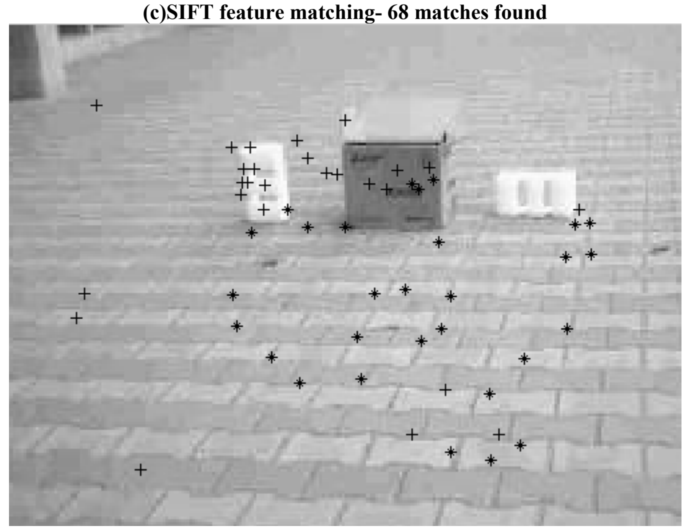
A set of outdoor images were collected from the campus of PEC University of Technology. Figure 5 - 6(a and b) shows some of the samples of stereo image pairs. The images were taken at different times and illumination conditions. Images of different ground planes i.e. tiled, cemented, grassed, etc. and artificial obstacles of different sizes were taken. Enough variation is kept to make the classification task challenging.



(a) Left Image **(b) Right Image**



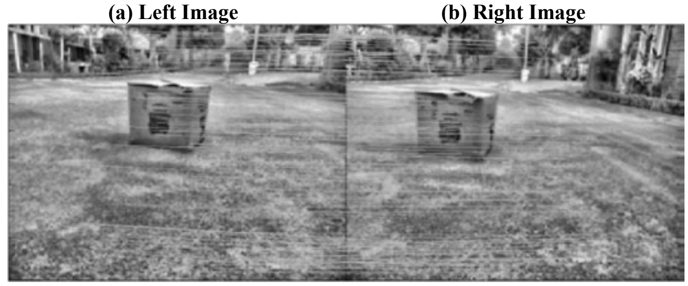
(c)SIFT feature matching- 68 matches found



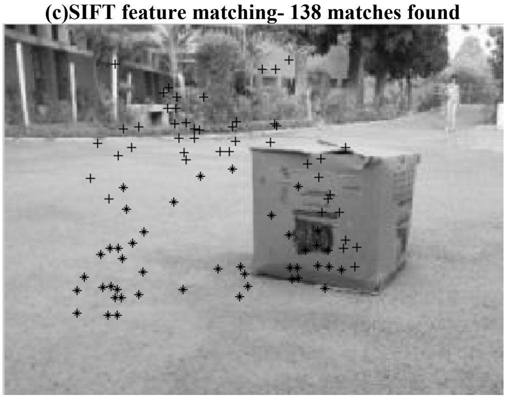
(d) Ground Plane Estimation - Ground Plane features ('*'), Non-ground plane features ('+')

Figure 5: "(a-B) stereo image pair-I, (c) sift matches in stereo images, (d) features classified as belonging to ground plane"

(a) Left Image **(b) Right Image**



(c)SIFT feature matching- 138 matches found



(d) Ground Plane Estimation - Ground Plane features ('*'), Non-ground plane features ('+' cross)

Figure 6: "(a-B) stereo image pair-II, (c) sift matches in stereo images, (d) features classfified as belonging to ground plane"

On the execution of the proposed approach, the results are collected in form of points marked in the image as points lying on the ground plane and another cluster of non-ground planes. The points which actually lie on ground plane and are marked as ground plane points are termed as true positives. Table 3 contains the result of execution of the proposed approach on the sample image pairs.

Image Pair→	I	II	III	IV	V
Ground Plane points found	30	56	116	28	67
Non-Ground Plane points found	31	52	114	29	68
True Positives	26	55	116	27	52
True Negatives	23	17	41	22	40
False Positives	8	35	73	7	28
False Negatives	4	1	0	1	15
True Positive Rate	86.6%	98.2%	100%	96.4%	77.6%

Table 3: Result of execution on Image Pair III

Figure 7 shows the ROC (Relative Operating Characteristic) curve. Points above the diagonal in the ROC curve clearly indicate that the proposed approach has good classification rate.

Experiments are also performed to determine the distance range, in which, the proposed approach has a good classification rate. At time of grabbing the images, the distance of the obstacles from the camera is noted. The analysis of the proposed approach with respect to distance is shown in form of a graph in Figure 8.

The experimental results indicate that the appropriate operating distance range for the proposed approach is 3 feet to 12 feet. If the obstacle is less than 2 feet away from the camera, the assumption that the lower 10% of the image is dominant which ground plane is violated. If the obstacle is placed very far away from the camera, the features of the obstacle planes are lost and it appears to be the same as the ground plane. Hence, the classification rate of the proposed approach is poor.

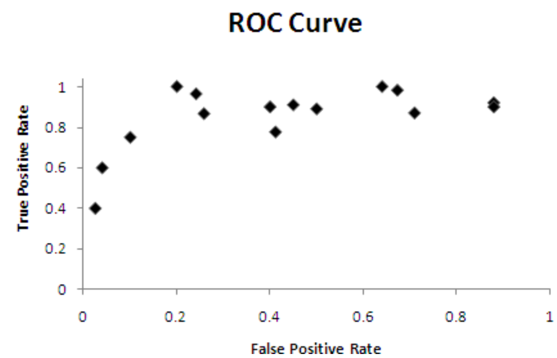


Figure 7: "ROC Curve"

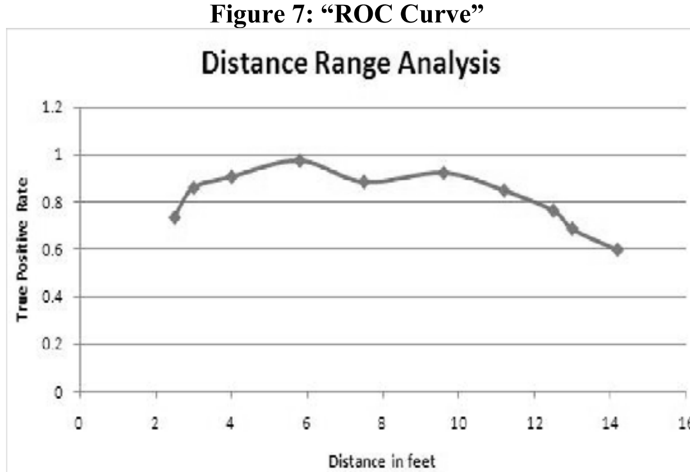


Figure 8: "True Positive rate Vs Distance graph"

6. CONCLUSIONS

The paper presents homography based approach for ground plane detection. The homography is very susceptible to the position of points used for its estimation. The paper presents a point selection algorithm which selects points based on fitness criteria for the accurate estimation of homography hypothesis. A homography estimate with minimum average distance error is used for classifying SIFT features as either belonging to ground plane or not. Generally, the feature points classified as lying on the ground plane are in the small region of the whole plane. This problem has been addressed by using image based segmentation using mean shift and normalized cuts. Experimental tests have been conducted to evaluate the performance of the proposed approach. The performed tests indicated that the proposed approach has good classification rate and operating distance range from 3 feet to 12 feet.

REFERENCES

[1]. World Health Organization. (Dec 2010). [Online] Available: http://www.who.int/

[2]. Vlad Coroama, Tarik Kapic, Feliz Rothenbacjer, "Improving the reality perception of visually impaired using pervasive computing", Advances in pervasive computing, OCG Press, 2004.

[3]. D. Dakopoulos, N.G. Bourbakis, "Wearable Obstacle Avoidance Electronic Travel Aids for the Blind: A Survey", IEEE Transactions on System, Man and Cybernetics, Part C, pp. 25, Jan 2010.

[4]. P.W. Nye, Ed., "A preliminary evaluation of the bionic instruments- Veterans administration C-4 laser cane", National Academy of Science, Final Rep. 1973.

[5]. G. Jansson, "The effect of the range of laser cane on detection of objects by the blind", Dep. of Psychol., Univ. of Uppsala, Sweden, Rep 211, 1975.

[6]. N. Pressey, "Mowat Sensor", Focus, vol. 11, no. 3, pp. 35-39, 1977.

[7]. J. Armgstrong, "Summary report of the research programme on electronic mobility aids", Dep. of Psychology, Univ. of Nottingham, Nottingham, England, 1973.

[8]. L. Kay, "Auditory Perception of Objects by blind persions using bioacoustic high resolution air sonar", JASA, Vol 107, pp 3266-3275, No. 6, June 2000.

[9]. S. Shoval, J. Borenstein and Y. Koren, "Mobile robot obstacle avoidance in a computerized travel aid for the blind", in the Proc. of IEEE Robot Automation Conference, San Diego, CA, pp. 2023-2029, May 8-13, 1994.

[10]. L.Russel, "Travel path sounder", In the proc. Of Rotterdam Mobility Conf., New York, 1965.

[11]. Allan G. Dodds, "The Nottingham Adjustment Scale: a validation study", International Journal of Rehabilitation Research, Vol. 16, No. 3, pp. 177-184, 1993.

[12]. Stephen Se, Michael Brady, "Ground plane estimation, error analysis and applications", Robotics and Autonomous systems, 39, pg 59-71, 2002.

[13]. J.J. Gibson, "The perception of the visual world", Houghton Mifflin, Boston, 1950.

[14]. N. Molton, S. Se, J.M. Brady, D. Lee, P. Probert, "A stereo vision-based aid for the visually impaired", Image and Vision Computing 16 (1998), pp. 251-263.

[15]. Scott Lenser and Manuela Veloso, "Visual Sonar: Fast obstacle avoidance using monocular vision", in IEEE Conference on Intelligent Robots and Systems, 2003.

[16]. J. Hoffmann, M.Jungel, M.Lotzsch, "Vision based system for goal-directed obstacle avoidance", in the 8th International Workshop on RoboCup, 2004.

[17]. L.M. Lorigo, R.A. Brooks, W.E.L. Grimsou, "Visually guided obstacle avoidance in unstructred environments", in the Proc. IEEE RSJ Int. Conf. Intelligent Robots and Systems, France, vol. 1, pp. 373-379, Sept. 1997.

[18]. M. Bertozzi, A. Broggi, "GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection", IEEE Transactions on Image Processing, pp. 62-81, 1998.

[19]. C. Daniel Herrera, "Ground surface estimation combining sterero data and optical flow", in the 10th International Symposium on Experimental Robotics, Brazil, 6-10, July 2006.

[20]. Nikolay Chumerin, Mark M. Van Hulle, " Ground plane estimation based on dense stereo disparity", in Proc. of the Fifth international conference on Neural Networks and Artificial Intelligence (ICNNAI'08),Minsk, Belarus, May 2-30, 2008.

[21]. Jun Zhao, Jayantha Katupitiya, James Ward, "Global Correlation Based Ground Plane Estimation using V-disparity Image", IEEE international conference on Robotics and Automation, Italy, pp. 529-534, 10-14 April, 2007.

[22]. R.Labayrade, D.Aubert, J.Tarel, "Real time obstacle detection on non-flat ground geometry through V-disparity representations", in Proc. of IEEE Intelligent Vehicle Symposium, 2002.

[23]. M.Betke, E. Haritaoglu, L.Davis, "Real time multiple vehicle detection and tracking from moving vehicle", Machine Vision and Applications, Springer, pp. 69-83, 2000.

[24]. Jin Zhou, Baoxin LI, "Homography based Ground Detection for a Mobile Robot Platform using a Single Camera", in the Proc. of IEEE International Conference on Robotics and Automation, pp. 4100-4105 May 2006.

[25]. D. Conrad, G.N.DeSouza, "Homography-based Ground Plane Detection for Mobile Robot Naviation Using a Modified EM Algorithm", in the Proc. of IEEE International Conference on Robotics and Automation, Alaska, May 3-8, 2010.

[26]. Muhammad Usman Keerio, "Obstacle Avoidance theough Teleoperation", BVICAM's International Journal of Information Technology, BIJIT, issue 2, Vol. 1, No. 2, 2009.

[27]. M Ali Reza, "Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement", Journal of VLSI Signal Processing Systems for Signal, Imahe and Video Technology, Vol. 38, pp. 35-44, August 2004.

[28]. David G. Lowe, "Distinctive image features from scale-invariant keypoints", International Journal of Computer Vision, 60(2), pp. 91-110, 2004.

[29]. M. Farenzena, A. Fusiello, R. Gherardi, R. Toldo, "Towards Unsupervised Reconstruction of Architectural Models", in Proc. of the Vision, Modeling and Visualization Workshop (VMV 2008), pp. 41-50, Konstanz, DE, Oct 8-10, 2008.

[30]. M.A. Fischler, R.C. Bolles, "Random Sample Consensus: A paradign model fitting with applications to image analysis and automated cartography", Communications of the ACM, 24(6), pp. 318-395, June 1981.

[31]. F.R. Hampel, P.J. Rousseeuw, E.M. Ronchetti, W.A. Stahel, "Roboust Statistics: the Approach based on Influence Functions", Wiley series in probabilty and mathematical statistics, John Wiley & Sons, 1986.

[32]. P.H. Torr, "An Assessment of Information Criteria for Motion Model Selection", in the Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp. 47-53, 1997.

[33]. Iwan Ulrich, "Appearance based obstacle detection with Monocular Color Vision", AAAI National Conference on Artificial Intelligence, Aug. 2000.

[34]. Nick Pears, Bojian Liang, "Ground Plane Segmentation for Mobile Robot Visual Navigation", in the Proc. of IEEE RSJ international Conference on Intelligent Robots and Systems, Maui, Hawaii, USA, Oct. 29- Nov. 03, 2001.

[35]. D. Santosh Kumar, C.V.Jawahar, "Robust homography based control for camera positioning in piecewise planar environment", in the Proc. of Indian Conference on Computer Vision, Graphics and Image Processing (IGVGIP), pp. 906-918, 2006.

[36]. Wenbing Tao, Hai Jin, Yimin Zhang, "Color Image Segmentation based on Mean Shift and Normalized Cuts", IEEE Transactions on Systems, Man, and Cybernetics, Part B, pp. 1382, Oct. 2007.