

An Advanced Stereo Vision Based Obstacle Detection with a Robust Shadow Removal Technique

Saeid Fazli, Hajar Mohammadi D., Payman Moallem

Abstract—This paper presents a robust method to detect obstacles in stereo images using shadow removal technique and color information. Stereo vision based obstacle detection is an algorithm that aims to detect and compute obstacle depth using stereo matching and disparity map. The proposed advanced method is divided into three phases, the first phase is detecting obstacles and removing shadows, the second one is matching and the last phase is depth computing. We propose a robust method for detecting obstacles in stereo images using a shadow removal technique based on color information in HIS space, at the first phase. In this paper we use Normalized Cross Correlation (NCC) function matching with a 5×5 window and prepare an empty matching table τ and start growing disparity components by drawing a seed s from S which is computed using canny edge detector, and adding it to τ . In this way we achieve higher performance than the previous works [2,17]. A fast stereo matching algorithm is proposed that visits only a small fraction of disparity space in order to find a semi-dense disparity map. It works by growing from a small set of correspondence seeds. The obstacle identified in phase one which appears in the disparity map of phase two enters to the third phase of depth computing. Finally, experimental results are presented to show the effectiveness of the proposed method.

Keywords—obstacle detection; stereo vision; shadow removal; color; stereo matching

I. INTRODUCTION

OBSTACLE detection is an important task for many mobile robot applications. Most mobile robots rely on range data for obstacle detection. Popular sensors for range-based obstacle detection systems include ultrasonic sensors, laser rangefinders, radar, stereo vision, optical flow and depth from focus. Because these sensors measure the distances from obstacles to the robot, they are inherently suited for the tasks of obstacle detection and obstacle avoidance.

Range sensors are also unable to distinguish between different types of ground surfaces. This is a problem especially in outdoors, where range sensors are usually unable to differentiate between the sidewalk pavement and adjacent flat grassy areas. Color provides more information than intensity alone. Compared to texture, color is a more local attribute and can thus be calculated much faster. Systems that solely rely on edge information can only be used in environments with texture-less floors [1]. This paper aims to research one of the computer vision's most important contributions to the navigation and obstacle detection for blind. Obstacle detection is defined as "the determination of whether a given space is free of obstacles for safe travel by an autonomous vehicle". Obstacle detection is one of the most renowned problems within the subfield of computer vision in terms of the amount of research it has attracted and the number of uses it has. Together with research into other subfields of artificial intelligence, obstacle detection is crucial in order to perform many basic operations for mobile robots such as avoidance and navigation. A good obstacle detection system must be capable of the following [2]:

- To detect obstacles on a given space in good time
- To detect and identify correct obstacles
- To identify and ignore ground features that may appear as obstacles

Stereo vision is an area of study in the field of machine vision that attempts to recreate the human vision system by using two or more 2D views of the same scene to derive 3D depth information about the scene. Depth information can be used to track moving objects in 3D space, gather distance information for scene features, or to construct a 3D spatial model of a scene.

As an emerging technology, stereo vision algorithms are constantly being revised and developed, and as we will be implementing a stereo vision system [3]. Stereo vision refers to the ability to infer information on the 3D structures and the distances of a scene from at least two images (left and right), taken from different viewpoints [4].

II. OBSTACLE DETECTION BASED COLOR SEGMENTATION AND STEREO VISION

Color segmentation, as the name implies, uses color to classify image areas as “obstacle” or “free-space”. The basic idea behind color segmentation for obstacle detection is that pixels in an image are classified as “obstacle” or “free-space” based on color. When operating in domains in which traversable areas are of relatively constant color, such as grass, color segmentation works well.

Each pixel in a color image consists of a 3-tuple, representing the amount of energy contained in the red, green, and blue bands. The images are not stored as standard (R, G, and B) tuples. Rather, they are first converted to a different color space, known as Hue-Saturation-Intensity, or HSI. This is a cylindrical space, in which the H and S components contain the color information, in the form of a standard color wheel. The Hue is the actual color, or the angle of the point in the cylinder, the Saturation is the “purity” of the color, and is the radial distance of the point. This space has the advantage that if we ignore the Value component, we get additional robustness to shadows and illumination changes, along with a reduction in feature dimensionality [5]. Color segmentation, although able to detect small obstacles and changes in color, is not sensitive to obstacle geometry, such as height [6]. Stereo vision is a powerful computer vision method for recovering 3-D scene structure, which works by comparing the differences between images taken by two cameras placed a short distance apart (like human eyes) to estimate depth. The fixed geometric relationship between the cameras simplifies depth estimation, making it a fast and relatively robust calculation. For our application, depth estimation is used to determine the ground plane, i.e. plane that the wheelchair is rolling on, and to locate obstacles, which are points in the scene that lie significantly above or below the ground plane. Stereo vision exploits the fact that a single point in a scene appears in slightly different locations in neighboring views of the scene. If the views are from two suitably aligned and calibrated cameras, with parallel lines of sight, a feature in the left image is horizontally shifted relative to the feature in the right image. This image shift, called the disparity, is directly related to the distance from the camera to the point

discuss in this paper, many alternative approaches exist for in the scene: distant points have small disparity, while nearby points have large disparity.

The following thought experiment illustrates this fact. If you alternately open and close your left and right eyes while looking at the sky, distant objects (such as stars or the sun) will appear in the same place in both eyes while nearby objects (such as your hand pointing towards the distant object) will appear in different locations in each eye [7]. A variety of work has been done on computer vision algorithms for finding negative obstacles using stereo depth information, primarily in the context of autonomous vehicles [8].

III. MATCHING ALGORITHM

Traditional area-based dense stereoscopic matching algorithms perform exhaustive search of the entire disparity space, i.e. they need to compute a correlation statistic for all putative correspondences [9]. Although efficient implementations for computing most of the commonly used statistics (SSD, SAD, NCC, SHD) are known [10], [2], this is still one of the most expensive phases of stereo matching. To avoid visiting the entire disparity space, algorithms were proposed that greedily grow corresponding patches from a given set of reliable seed correspondences. Such algorithms assume that neighboring pixels have similar disparity, not exceeding disparity gradient limit [11] or a similar constraint. The principle of growing a solution from initial seeds had long been known in segmentation [12].

To simplify the description of the algorithm, we assume a pair of horizontally rectified stereo images is used. Generalization to unrectified images is possible but it will not be discussed in the present paper. We therefore assume we are working with matching table. It represents a 3D discretized disparity space in which each element (x, x', y) denotes a possible correspondence $(x, y) \leftrightarrow (x', y)$, x is horizontal axes of left image, x' is horizontal axes of right image and y is vertical axes of both images. Each matching table element (x, x', y) may be associated with some parameters modeling relative distortion of image neighborhoods. The parameters may be updated during the growth process to accommodate to the slant of the 3D surface, as in [13]-[16].

This is important in wide baseline stereo. For simplicity, we omit the distortion model here. Suppose we are given an unsorted list of disparity seeds S . Each seed is a point in disparity space, $s = (x, x', y)$. Its neighborhood $N(s)$ in disparity space consists of 16 point constructed from four

sub-sets $N_1(s) \cup N_2(s) \cup N_3(s) \cup N_4(s)$, see Fig. 1 where we use the same colors for:

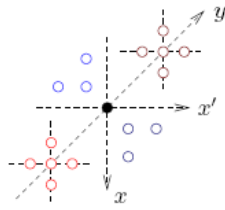


Fig. 1. Disparity space neighborhood used in this paper.

$$N1(s) = \{(x-1, x'-1, y), (x-2, x'-1, y), (x-1, x'-2, y)\}$$

$$N2(s) = \{(x+1, x'+1, y), (x+2, x'+1, y), (x+1, x'+2, y)\}$$

$$N3(s) = \{(x, x', y-1), (x \pm 1, x', y-1), (x, x' \pm 1, y-1)\}$$

$$N4(s) = \{(x, x', y+1), (x \pm 1, x', y+1), (x, x' \pm 1, y+1)\}$$

The neighborhood is selected so as to limit the magnitude of disparity gradient to unity and to improve the ability to follow a disparity component even if the image similarity peak falls in between pixels in the matching table. Assuming similarity is computed from small image windows around pixels (u, v) and (u', v) by e.g. the normalized cross-correlation, we prepare an empty matching table τ and start growing disparity components by drawing an arbitrary seed s from S , adding it to τ , individually selecting the best-similarity neighbors q_i over its four subneighborhoods $N_i(s)$:

$$q_i = (u, u', v) = \underset{(x, x', y) \in N_i(s)}{\arg \max} \text{simil}(x, x', y)(1)$$

and putting these neighbors q_i to the seed list if their interimage similarity exceeds a threshold. Hence, up to four new seeds are created. If we draw a seed from the list S that is already a member of the matching table, then we discard it. The growth must stop in a finite number of steps by exhausting the list S . The output from the growth phase is a partially filled matching table whose connected regions in 3D represent disparity components grown around the initial seeds. Note that disparity components obtained this way are nothing more than contiguous segments in disparity space. In the extreme case, the entire disparity space is filled by a single component grown from the first seed. Obviously, such growth is not a very efficient way of selecting high-similarity tentative matches. Instead, we see it only as an elementary mechanism for traveling in disparity space. This phase has been introduced just to show correctness of the

entire proposed algorithm which will be described shortly. Note, finally, that the order of selecting the seeds from the list S is arbitrary [17]-[19]. The matching task is done by solving a graph theoretic problem known as the maximum strict sub-kernel (SSK) [20], [21].

Strict sub-kernel is a general notion valid for oriented graphs, in which the graph structure represents the structure of the underlying problem (the structure of constraints) and the orientation represents evidence (data) [20]. The notion of SSK is related to the well-known Stable Marriage and Stable Roommates Problems [22]. The paper [17] uses S matrix as random but we use canny (or sobel) edge detector. Therefore the method achieves higher performance as shown in the experimental results.

IV. PROPOSED METHOD

The proposed method is divided into three phases, the first one is detecting obstacles and removing shadows, the second phase is matching and the last one is depth computing. We propose a robust method for detecting obstacles in stereo images using shadow removal technique based on color information in HIS space. Our work focuses on a different domain, in which the features of interest are difficult to detect from depth information alone, because the depth changes that characterize these obstacles may be small relative to the distances at which they are viewed and may be swamped by noise in the depth information estimated from computer vision stereo algorithms.

The first phase consists of four steps as follow: In the first step, color input image RGB values are transformed into the HSI (hue, saturation, and intensity) color space. Because color information is very noisy at low intensity, we only assign valid values to hue and saturation if the corresponding intensity is above a minimum value. Similarly, because hue is meaningless at low saturation, hue is only assigned a valid value if the corresponding saturation is above another minimum value. An appealing attribute of the HSI model is that it separates the color information into intensity and color components. As a result, the hue and saturation bands are less sensitive to illumination changes than the intensity band.

In the second step, since not every pixel is assigned a valid hue value, the hue matrix of RGB image is filtered with a 7×7 Median filter to reduce the noise level.

In the third step, a rectangular area in front of the mobile robot or blind person is used for reference in hue and intensity, because the mobile robot or blind person walk at first in ground that is surely background in images. Then the

average of reference area is computed for both hue and intensity.

In the fourth step, all pixels of the input image are compared to average of the hue and the intensity. A pixel is classified as shadow if both of the two following conditions are satisfied:

I) The difference of hue value at the pixel's hue value and the average of reference area in hue is below the hue threshold.

II) The intensity value at the pixel's intensity value is above the average of reference area in intensity.

If first condition and following condition are true the pixel is satisfied as ground:

II) The intensity of pixel is between minimum and maximum of area reference in intensity matrix.

Otherwise the pixel is satisfied as an obstacle. The hue threshold is 20% of hue difference in rectangular reference area.

The global algorithms for matching formulate the problem of the disparity computation as an energy-minimizing problem and the smoothness assumption of the disparity map is usually made, there is some kind of function matching for stereo matching such as; Sum of Squared Difference (SSD), Sum of Absolute Difference (SAD), Normalized Cross Correlation (NCC) and Sum of Hamming Distance (SHD). The NCC function is function matching with demanded accuracy which is calculated by above equation:

$$NCC(f, g) = \frac{(f - \bar{f}) \cdot (g - \bar{g})}{|f - \bar{f}| |g - \bar{g}|} \quad (2)$$

In this paper we use NCC function matching on 5×5 window as image similarity statistic in all experimental results. We prepare an empty matching table τ and start growing disparity components by drawing a seed s from S which is computed by using *canny* or *sobel* edge detector, adding it to τ . In this way we have higher performance than previous works.

A fast stereo matching algorithm is proposed that visits only a small fraction of disparity space in order to find a semi-dense disparity map, for second phase. It works by growing from a small set of correspondence seeds. Unlike in known seed-growing algorithms, it guarantees matching accuracy and correctness, even in the presence of repetitive patterns. The quality of correspondence seeds influences computing time, not the quality of the final disparity map. Finally the third phase, depth of obstacle is computed as follows:

$$Z(x, y) = \frac{f \cdot B}{d(x, y)} \quad (3)$$

f is focal length, B is base line (distance of cameras center), $d(x, y)$ is disparity value in x row and y column. The participation of both phase 1 and phase 2 output is input of third phase.

The block diagram of the proposed method is shown in fig. 2.

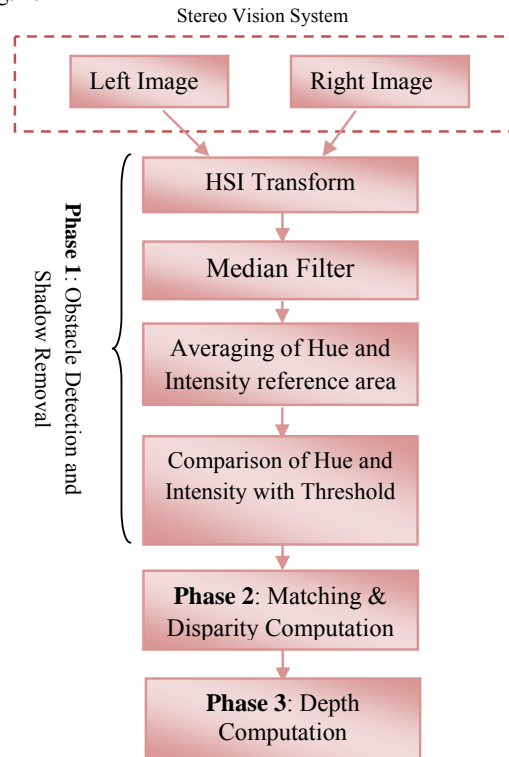


Fig. 2. The block diagram of proposed method

V. EXPERIMENTAL RESULTS AND CONCLUSIONS

The experimental results that are shown have been captured by uncalibrated cameras. Therefore a disparity map, computed by traditional methods results in many errors and low accuracy. The proposed method outperforms the previous works [2], [17]. Figures 3-6 show the outputs of phase 2 without applying the shadow removal technique. As can be seen the shadows are detected as obstacles. After applying the proposed shadow removal technique, although there are some shadows in the original image (fig. 7), but the algorithm is able to remove them and keep just real and correct obstacles. Figures 7-14 show the final results. The matching time is about one sec in the MATLAB running on a Pentium 4 PC, which is faster than the one of random seed which is about 3.3 sec on the same computer.

As the cameras are uncalibrated, we need one stage to compute the similarity between left and right images. This similarity is shown in fig. 15. The results shown here present that the proposed method is quite fast and robust against shadows.



Fig. 3. Original left and right images

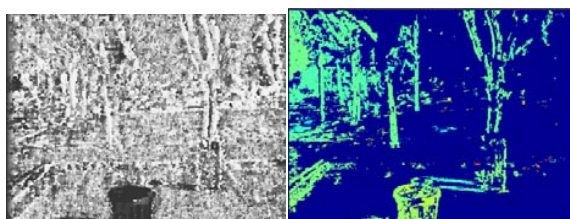


Fig. 4 Disparity map of last figure computed with SSD method (left), and the proposed method using sobel edge detector (only Phase 2, right)



Fig. 5. Edge detection by canny operator (only Phase 2)

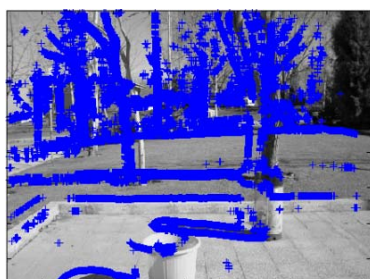


Fig. 6. The points marked in blue are the elements of seed matrix S (only Phase 2)

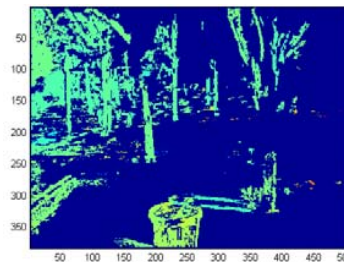


Fig. 7. Disparity map computed with the proposed algorithm using canny operator (only Phase 2)



Fig. 8, Original left and right images



Fig. 9. Saturation of Fig. 8



Fig. 10. Hue of Fig. 8



Fig. 11. Rectangular reference area



Fig. 12. Detected obstacles

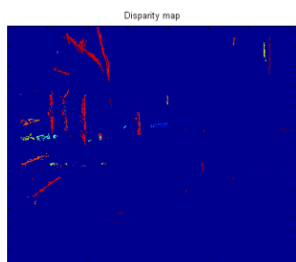


Fig. 13. Disparity map without applying the shadow removal technique

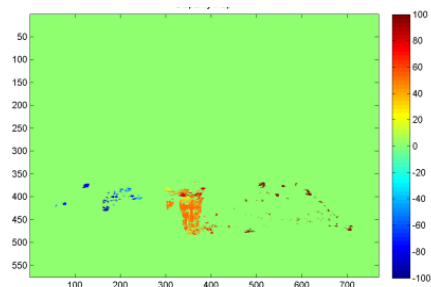


Fig. 14. Disparity map with applying the shadow removal technique



Fig. 15. Similarity of left and right images

REFERENCES

- [1] I. Ulrich, I. Nourbakhsh, "Appearance-Based Obstacle Detection with Monocular Color Vision", AAAI National Conference on Artificial Intelligence, Austin, TX, July/August 2000.
- [2] S. Fazli, H. Mohammadi Dehnavi, and P. Moallem, "A Robust Obstacle Detection Method in Highly Textured Environments using Stereo Vision", 2nd international conference on machine vision (IEEE), p 97-100, 2009.
- [3] P. Munro, A. P. Gerdelan, "Stereo Vision Computer Depth Perception", Country United States City University Park Country

Code US Post code 16802 Number of Publications 1,358,061 Number of Authors 1,096,162 Date of last added publication 2009.

- [4] P. Foggia, Jean-Michel Jolion, A. Limongiello, and M. Vento, "Stereo Vision for Obstacle Detection: a Graph-Based Approach", 6th IAPR – TC-15 Workshop on Graph-based Representations in Pattern Recognition (Gbr 2007), Alicante June 11-13, 2007.
- [5] P. H. Batavia, S. Singh, "Obstacle Detection Using Adaptive Color Segmentation and Color Stereo Homography", IEEE International Conference on Robotics and Automation, May, 2001.
- [6] Y. Chao, Z. Changan, "Obstacle Detection Using Adaptive Color Segmentation and Planar Projection Stereopsis for Mobile Robots", IEEE International Conference on Robotics, Intelligent Systems and Signal Processing, October 2003.
- [7] J. Coughlan, H. Shen, "TERRAIN ANALYSIS FOR BLIND WHEELCHAIR USERS: COMPUTER VISION ALGORITHMS FOR FINDING CURBS AND OTHER NEGATIVE OBSTACLES", Conference & Workshop on Assistive Technologies for People with Vision & Hearing Impairments Assistive Technology for All Ages CVHI 2007.
- [8] P., Bellutta, R., Manduchi, L., Matthies, K. Owens, and A. Rankin, Terrain Perception for DEMO III. Intelligent Vehicle Symposium 2000.
- [9] D. Sharstein, R. Szeliski, and R. Zabih. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. IJCV, 47(1):7–42, 2002.
- [10] P, Moallem and K, Faez, "Effective Parameters in Search Space Reduction Used in a Fast Edge-Based Stereo Matching", Journal of Circuits, Systems, and Computers, 14(2): 249-266, 2005.
- [11] S. B. Pollard, J. E. W. Mayhew, and J. P. Frisby. PMF: A stereo correspondence algorithm using a disparity gradient constraint. Perception, 14:449–470, 1985.
- [12] R. M. Haralick and L. G. Shapiro. Image segmentation techniques. CVGIP, 29:100–132, 1985.
- [13] Z. Megyesi, G. K'os, and D. Chetverikov. Dense 3D reconstruction from images by normal aided matching. Machine Graphics and Vision, 15:3–28, 2006.
- [14] G. P. Otto and T. K. W. Chau. 'Region-growing' algorithm for matching of terrain images. IVC, 7(2):83–94, 1989.
- [15] M. A. O'Neill and M. I. Denos. Practical approach to the stereo matching of urban imagery. IVC, 10(2):89–98, 1992.
- [16] T. Kim and J. Muller. Automated urban area building extraction from high resolution stereo imagery. IVC, 14:115–130, 1996.
- [17] Jan Cech, Radim Sara. Efficient Sampling of Disparity Space for Fast and Accurate Matching. In *Proc. BenCOS Workshop CVPR*, 2007.
- [18] Radim Sara. Finding the largest unambiguous component of stereo matching. In *Proc. ECCV*, pp. 900-14, 2002.
- [19] Radim Sara. Robust Correspondence Recognition for Computer Vision. In *Proc COMPSTAT, 17th Conference of IASC-ERS Roma*, Italy, 2006.
- [20] R. Sara. Finding the largest unambiguous component of stereo matching. In *Proc ECCV*, pp. 900–914, 2002.
- [21] R. Sara. Robust correspondence recognition for computer vision. In *Proc COMPSTAT*, pp. 119–131. Physica-Verlag, 2006.
- [22] D. Gusfield and R.W. Irving. The Stable Marriage Problem: Structure and Algorithms. The MIT Press, 1989.