Feature-Based 3-D Surface Reconstruction Directed by Grid Space Projection

Mohammed T. El-Sonni Wesam M. Dawoud

College of Computing Engineering
Arab Academy of Science and Technology and Maritime Transport
P. O. BOX 1029, Alexandria, Egypt

Abstract: - In this paper we describe a voxel-based 3-D reconstruction algorithm from multiple calibrated camera views. Unlike image-based algorithms, this algorithm is capable of detecting occlusion explicitly, and recovering the conventional Stereo Algorithms limitations; the algorithm is extendable to reconstruct the full surface without any restrictions on the cameras distribution. Because of using stable features at consistency checking the mismatching probability is decreased. The Grid Space is traced one time only; hence, there is no any additional computation or memory consuming. In spite of that, experimental results on both real and synthetic images show the algorithm efficiency and time preserving.

Key-words:- Stereo Vision**,** 3-D reconstruction, Grid-Space, Feature-Based

1 Introduction

Surface reconstruction algorithms have many practical applications from terrain mapping and industrial automation to Virtual Reality and real-time human-computer interaction. The 3-D reconstruction from multiple camera views can be distinguished into two classes of algorithms. The First class is the standard Stereo algorithms (image-based); they try to represent a scene from image pairs. A challenging is matching features between these images robustly. These methods typically employ normalized cross correlation along epipolar lines. Several authors have pursued volumetric representation to assist in this task, typically with image coordinates for two axis and disparity hypothesis being the third axis. The earliest example is given by [1] who develops a relaxation network that enforce uniqueness and continuity constraint by introducing inhibitory and excitatory connections between voxels representing disparity hypothesis.

Marr at [1] uses discrete match values and a 2-D local support area possible due to memory and processing constraint, so his experiments was restricted to synthesized images, Zitnick [2] replace the 2-D local support area with 3-D local support area and construct a new update function, his algorithm can be run on both real and synthesized images. Zhang [3] extends the original cooperative algorithms [1] and [2] in two ways. First, he designed a method of adjusting the initial matching score volume to guarantee that correct matches have high matching scores. Second, he developed a scheme for choosing local support areas by enforcing the image segmentation information. As a result, the foreground fattening errors are drastically reduced.

Numerous stereo matching algorithms, including local matching (e.g., [4], [5]), global optimization (e.g., [6], [7], [8], [9]), dynamic programming (e.g., [10],[11]) have been proposed over the past decades, all these methods are often limited for several reasons, First , input views can only be separated by a limited distance (baseline) , for correlation to be effective. Second, the result of stereo reconstruction is at best a 2½D reconstruction. Third, occlusion process is difficult to model image space.

The Second class of 3-D reconstruction algorithms is object space (3-D space) algorithms, in 3-D space it is easer to reason about occlusion relationships, as well as identify corresponding regions for correlation in image space. Seitz [12] proposed a voxel-based algorithm assumes that, the object is contained within a volume , this volume is traced to determine the object surface, but the main disadvantages of this algorithm is the constraint introduced on the possible camera views, this constraint restrict the type of scenes that can be reconstructed, This problem has been addressed in [13] by a multi plane sweep technique along x, y, z coordinates, and in both positive and negative directions, Culbertson at [14] try to eliminate the multiple sweeping at [13] by proposing algorithm called Generalized Voxel Coloring (GVC) , where a layered depth image (LDI) data structure has been utilized so that every pixel is associated with a linked list of voxels that project over it , sorted in depth order. The disadvantages with LDI implementation is the memory consuming in case of large Grid Space.

The remainder of this paper is structured as follows: Section II will present the Proposed Algorithm, Section III demonstrates by experiments the effectiveness of the proposed algorithm compared with the others algorithms, and finally, Section IV will conclude this work.

2- Surface Reconstruction

2.1 Volume Initialization

The first step of the proposed reconstruction algorithm is to define a volume in the reference coordinate system that encloses the 3-D object to be reconstructed. The volume extensions are determined from the calibrated camera parameters and its surface represents a conservative bounding box of the object. The volume is discretized in all three dimensions leading to an array of voxels (Grid Space), where the position of each voxel in the 3-D space is defined by its indices (*l, m, n*).

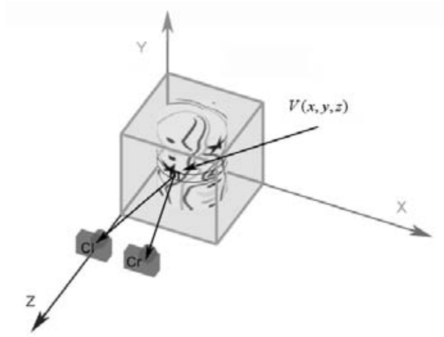


Figure 1 Object at Grid Space

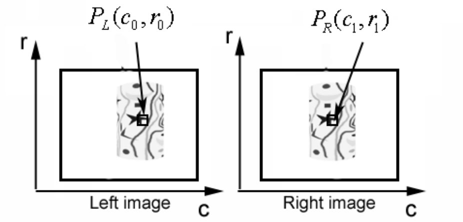


Figure 2 The projection of Voxel P into both cameras

Let (X_i, Y_i) being the pixel position of the perspective of the voxel (x_l, y_m, z_k) into the i^{th} camera view, then the projection of the voxel for view i is obtained as

$$X_i = f_x \frac{x_{li}}{z_{ni}} \text{ , } Y_i = f_y \frac{y_{mi}}{z_{ni}} \quad (1)$$

with

$$(x_{li}, y_{mi}, z_{ni})^T = R_i(x_l, y_m, z_n)^T + T_i \quad (2)$$

R_i and T_i are the camera rotation and translation at position i with respect to reference coordinate system. The parameters f_x and f_y describe the camera geometry and the scaling that relates pixel coordinate to world coordinates.

2.2 Voxels Classification

In the second step of the proposed reconstruction algorithm the voxels are classified into *Useful* and *Unuseful voxels*, *Useful voxels* are that lie on edges caused by sharp changes in the surface color or orientation in the x direction, while the *Unuseful voxels* are voxels lie on the texturless areas of the surface, using this voxels at reconstruction increase the probability of mismatching, so it is ignored during the Grid Space tracing. The algorithm flow chart become as the following:

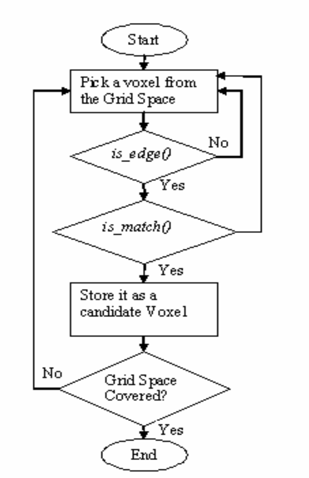


Figure 3 Flowchart of Voxels Classification process

is_edge() tests if the projection of the traced voxel contains an edge , if it is satisfied , algorithm continue to the next test , *is_match()* return a true when the (3) is true,

$$\frac{|I_L avg - I_R avg|}{|I_L avg + I_R avg|} < \Theta \quad (3)$$

where $I_L avg$ and $I_R avg$ is the average of the illumination of the voxel projection window at the left and right images consecutively , and Θ is a threshold predefined by experiments. With the colored images (3) will be applied to each color separately, so for a voxel to be candidate its projection into the images must satisfy:

$$\frac{|R_L avg - R_R avg|}{|R_L avg + R_R avg|} < \Theta_R \ , \ \frac{|G_L avg - G_R avg|}{|G_L avg + G_R avg|} < \Theta_G \ , \ \frac{|B_L avg - B_R avg|}{|B_L avg + B_R avg|} < \Theta_B \quad (4)$$

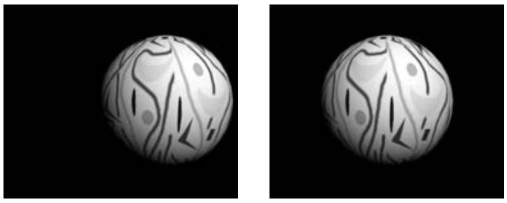
where Θ_R, Θ_G and Θ_B is a threshold for each color channel.

The normalization of the color components is used to increase the robustness of the reconstruction algorithm with respect to the varying illumination conditions.

As shown above, the Grid Space never stored at memory, it is trace once and the candidate voxels are stored in 2-D buffer, thus the reaming processes will run in 2-D space only.

2.3 Surface interpolation:

To show the output of the second step assume the following stereo images of a textured sphere,



(a) (b)

Figure 4 stereo images taken for a textured ball at 3D stereo max

The algorithm is applied on the middle slice of the Grid space, as shown in the figure 5, only the edges point which is candidate to be on the surface stays, the others are pruned.

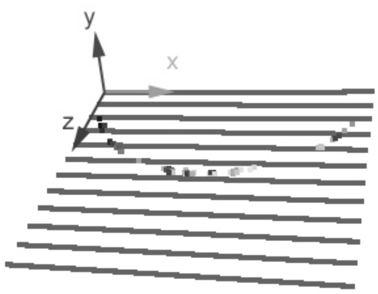


Figure 5 The contents of the projection buffer when the middle slice of the ball is scanned, only the edges voxels appear.

The candidate voxels are scattered on the surface, to reconstruct the surface, they must connected in some how.

Projection Buffer has the solution; the most important characteristic of the Projection Buffer is that, the adjacent cells in the projection buffer actually are a projection of adjacent voxels on the object surface.

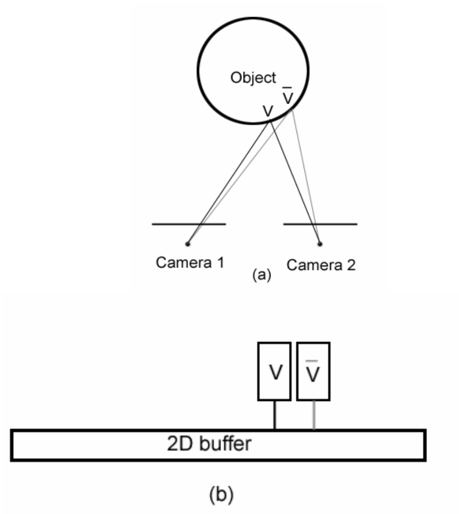
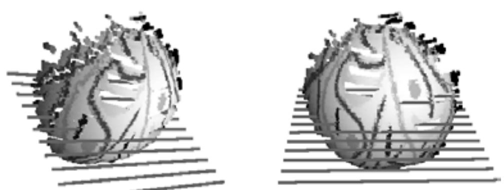


Figure 6 (a) two adjacent points on the surface projected into stereo cameras, (b) the adjacent voxels is stored at the Projection Buffer.

Voxel V is connected with $\bar{V}$ if the Euclidian distance between them on the surface less than a threshold.

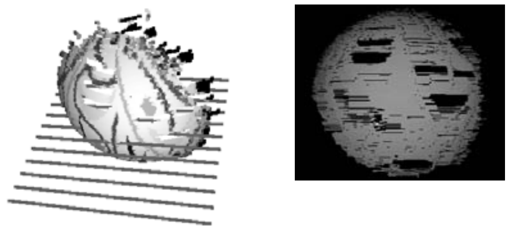
$$r = \sqrt{(\bar{x} - x)^2 + (\bar{y} - y)^2 + (\bar{z} - z)^2} < thresold \text{ , } (3)$$

where (x, y, z) is the source point and $(\bar{x}, \bar{y}, \bar{z})$ is the destination point, but a question arise here, how does the value of this threshold is chosen? The small value cause a coarse discontinuity of the surface, while a big value may assume a different surfaces as a one surface. Hence, the best value of the *threshold* can be calculated during the Grid Space tracing phase by taken the average of the distances between the adjacent candidate points on the surface.



(b)

(a)

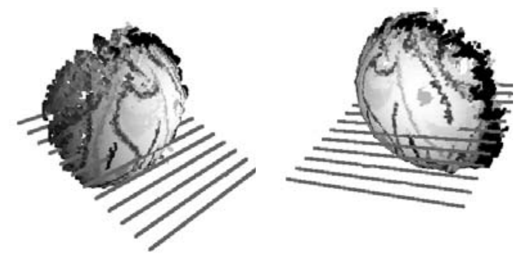


(c)

(d)

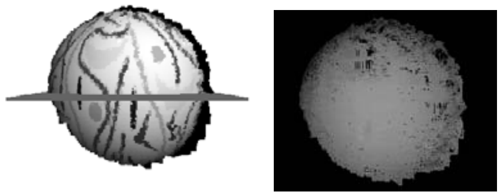
Figure 7 (a) to (c) are the plotting of the contents of the projection buffer into the space (d) depth map, which also show the last distribution of the surface voxels at the projection buffer

The discontinuities are clear at figure 7, the cause is the wide areas that doesn't have a noticeable color or surface changes , additional connecting iteration in the y direction will reduce these discontinuity, see the next figure,



(a)

(b)



(c)

(d)

Figure 8 (a) to (c) are the plotting of the contents of the projection buffer into the space after connecting in the y direction (d) is the depth map

2.4 Full surface reconstruction:

To extend this algorithm to reconstruct the full object surface, the object must captured from different scenes which recover its entire surface, each adjacent camera positions have their Projection Buffer, after applying the proposed Stereo algorithm on each of them , the full surface of the object is reconstructed,

Figure 9 The full surface reconstruction by 18 surrounding images

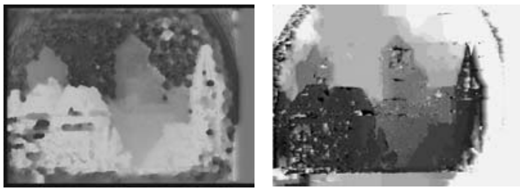
3 Experiments

To show the generalization of the algorithm, it was necessary to apply it on real images created by others, then compare the results with that given by well known stereo algorithms, the datasets is created by Carnegie Million University (CMU), and the stereo algorithms compared with are [Zitnick 00] and [Roy 98], these algorithms are selected because they offer a global solution to the stereo matching problem.

Datasets #1:

The left image is number five from the 11 images sequence provided by CMU as experiment1 while the right image is number three from the sequence,

(a) (b)



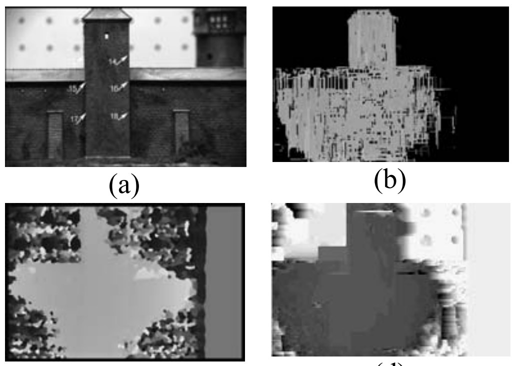
(c) (d)

Figure 10 (a) The right Image (b)the depth map constructed by the proposed algorithm, (c) the disparity map constructed by Zitnick's algorithm and finally (d) is the disparity map constructed by Roy's algorithm.

As shown in figure 10(b) the proposed algorithm discover that the tower at the right is separated from lower tower on it's left while the other algorithms at 10(c),(d) continue with the lower tower until they reach the tower on the right.

Datasets #2:

This images will test the algorithms in case of low features images, The depth maps at figure 11 show the failing of both Zitnick's and Roy's algorithms at the top of the tower

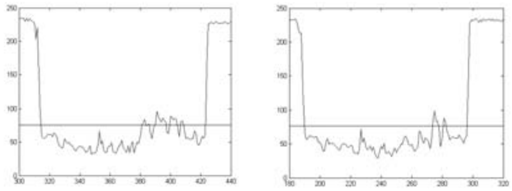


(a) (b) (c) (d)

Figure 11 (a) The right Image (b)the depth map constructed by the proposed algorithm, (c) the disparity map constructed by Zitnick's algorithm and finally (d) is the disparity map constructed by Roy's algorithm.

To understand the source of the problem, figure 12 shows part from the scanline 117 of the left and right images.

Each intensity at left image can find its corresponding at the right image until x=380.



(a) (b)

Figure 12 (a) part from the scanline 117 of the left image , (b) the scanline at the right image

At that point the intensity changes at the left image is differ than that at the right images , This problems did not affect the proposed approach , the proposed approach depends on edges which are more stable , figure 11 (b) show how the top of the tower is properly reconstructed.

The table shows the running time of the proposed algorithm compared of that of [6] and [2] algorithms on Intel Pentium III processor, 799MHz, 384 MB of RAM, the tested images are 576x384 pixel, and the measured time is in minutes:

Dataset #	Algorithm	time (minutes)
1	Our Algorithm	2.5
	[Zitnick 00]	3.2
	[Roy 98]	4.45
2	Our Algorithm	2.5
	[Zitnick 00]	4.25
	[Roy 98]	5.6
3	Our Algorithm	2.12
	[Zitnick 00]	3.25
	[Roy 98]	4.9

Table 1 The running time of the proposed algorithm compared with that of [6] and [2] Algorithms.

It is seen from the table1 that, the running time of the proposed approach, always is the minimum.

4 Conclusions & Future works

The proposed approach contributes as a Stereo Algorithm extendable to reconstruct the full surface; its efficiency has shown by experiments. The range of applications recovered by this approach increase if it became able to treat dynamic objects, this require online and parallel processing, actually each part of the space can be processed separately on a different processor.

5 References

[1] D. Marr and T. Poggio, "Cooperative Computation of Stereo Disparity," Science, Vol. 194, No. 4262, Oct. 1976, pp. 283-287.

[2] C. Lawrence Zitnick and Takeo Kanade, "A cooperative algorithm for stereo matching and occlusion detection" IEEE Trans, July 2000, PAMI, 22(7):675–684.

[3] Ye Zhang and Chandra Kambhamettu, "Stereo matching with segmentation-based cooperation" In Proc. ECCV, 2002 ,volume 2, pp. 556–571.

[4] T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window: Theory and experiment. IEEE Transactions on Pattern Analysis and Machine ntelligence, 16(9):920–932, September 1994.

[5] P. Fua. A parallel stereo algorithm that produces dense depth maps and preserves image features. Machine Vision and Applications, 6(1):35–49, 1993.

[6] S. Roy and I. Cox. "A maximum-flow formulation of the N-camera stereo correspondence problem" In Proc. ICCV, 1998, pp. 492–499.

[7] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. In Proceedings of IEEE Computer Society International Conference on Computer Vision, pages 377–384, 1999.

[8] M. Agrawal, and L. Davis, "Window-based discontinuity preserving stereo". CVPR 2004.

[9] L. Hong and G. Chen. "Segment-based stereo matching using graph cuts" CVPR 2004.

[10] I. J. Cox, Hingorani, S. B. Rao, and B. M. Maggs, "A Maximum Likelihood Stereo Algorithm," Computer Vision and Image Understanding, 1996, vol. 63, pp. 542-567.

[11] S. Birchfield and C. Tomasi, “Depth Discontinuities by Pixel-to-Pixel Stereo,” Proc. IEEE Int’l Conf. Computer Vision, 1998, pp. 1073-1080.

[12] S. Seitz and C. Dyer, “Photorealistic Scene Reconstruction by Voxel Coloring,” Proceedings of the IEEE Conference onComputer Vision and Pattern Recognition, June 1997, pp. 1067-1073.

[13] K. N. Kutulakos and S. M. Seitz, “A Theory of Shape by Space Carving,” International Journal of Computer Vision, Vol. 38, No. 3, July 2000, pp. 199-218.

[14] W. Culbertson, T. Malzbender, and G. Slabaugh. “Generalised voxel coloring” volume 1883, Springer Verlag, 2000 pp. 100-115.