International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

PVHL-based Local 3D Map Building

Huahua Chen[1] **Xueyi Ye**

College of Communication Engineering
Hangzhou Dianzi University
Hangzhou 310018,China

Abstract: 3D map building is an important task of autonomous land vehicle(ALV) for obstacle detection, path planning. This paper analyzes the defects of the traditional 3D map building method, and then proposes a novel method based on projection of virtual height line(PVHL). The method projects virtual height line(VHL) into the stereo vision system by transforming the world coordinate of each point on VHL into its image coordinate, and generates two projection lines(PLs) in two images. On the two PLs, only the correspondence that is the projection of the intersection point of VHL and the surface of the scene has the maximum similarity metric. Then finding the height of the scene point is equivalent to proving whether the real correspondence of VHL has the maximum similarity metric or not. Experiments show that the proposed method is effective. And compared with that of the traditional, the results show that the proposed avoids unnecessary computation time and has better performance than the traditional, and can build the interested region only.

Keywords: **Virtual height line(VHL); Stereo matching; 3D map building**.

1. Introduction

Recently, there has been a great interest in using autonomous land vehicle(ALV) to build 3D maps of the environment, and many studies about map building exist in the literature[1-4]. In many respects, ALV moves in an unknown environment and it will use the surrounding information to build a map which is used to for location and navigation. Map building is a precondition of the true autonomous navigation in essence.[4].

The ALV[1] requires a real-time planner using the local map built from the data of the radar which can detect and update obstacles with a very high speed. But radars are active sensors and are easier to be found by the enemy in some special scene, such as battlefield. Besides, radars are seldom used in cross-country environment because of its lower angle resolution. In recent years, as high performance processors appear, the same detection performance with laser radar at a distance less than 100 feet can be implemented by using a real-time stereo vision algorithms. In addition, a stereo vision system can measure distances faster and more accurately by adopting passive sensors which can work without radiation sources, so it has very good concealment. Furthermore, the system uses non-scanning imaging to measure distances, so it is favorable for ALV moving under harsh terrain.

The traditional 3D map building method usually relies on dense disparity map obtained by stereo matching[1]. Although stereo matching algorithms have achieved remarkable success[5], [6] and many

[1] Huahua Chen
Email address: science_ant@hotmail.com

fast methods of global stereo matching are proposed [6], [7] in recent years, they all require the very high performance computing ability, and even demand the high performance GPU to aid computation [7]. However, most onboard computing ability or power supply of small robots is limited by energy, weight, size, etc. Therefore, a fast and robust map building method with low computation complexity is still of interest in many applications.

In past years, 3D reconstruction methods without disparity map were proposed. I.S. Kweon et al. proposed the locus method which follows the image formation process[8],[9]. It projects a hypothetical vertical line into a range image to create a locus, and then computes the intersection of depth surface with locus curve in image space rather than Cartesian space. This method avoids dense stereo matching, but its performance heavily depends on the precision of the range image.

In this paper, a novel 3D map building method is proposed. Firstly, we introduce virtual height line(VHL) into the scene, then project VHL into the two images of the stereo vision system by changing world coordinate of each point on VHL into its image coordinate, and generates two projection lines(PLs) in stereo vision system. The intersection point of VHL and the surface of the scene, which is the only existent and visible point on VHL, exists the real correspondence that has the maximum similarity metric on the two PLs. So finding the height of the scene point is equivalent to proving whether the projected correspondence has the maximum similarity metric or not. This method avoids unnecessary computation time. Experiments show that the proposed method is effective.

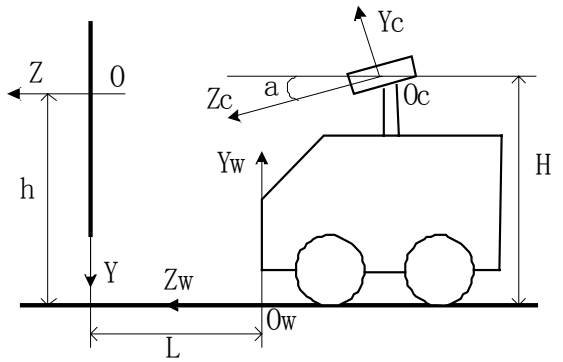


Figure 1 Relationship between the C-CS and the W-CS.

2. Traditional 3D map building method

Traditional 3D map building method usually relies on dense disparity map. A 3D position of each scene point can be obtained from its disparity and parameters of the stereo vision system. The 3D map of a scene is firstly computed by triangulation in camera coordinate system, as shown in equation (1).

$$\begin{cases} Xc = \dfrac{b\times(U-U_0)}{d} \\ Yc = \dfrac{b\times(V-V_0)}{d} \\ Zc = \dfrac{b\times f}{d} \end{cases} \tag{1}$$

where (X_c, Y_c, Z_c) is the camera coordinate system, and Z_c is the optical axis. (U_0, V_0) is the principal point, b is the length of the base-line, f is the focal length, and d is the disparity value of the pixel (U,V).

Figure1 shows the relationship between the camera coordinate system(C-CS) $O_cX_cY_cZ_c$ and the world coordinate system(W-CS) $O_wX_wY_wZ_w$. O_cX_c and O_wX_w are perpendicular to the paper outwards. The relationship between $O_cX_cY_cZ_c$ and $O_wX_wY_wZ_w$ can be described as equation (2).

$$\begin{cases} X_c = X_w\cos\gamma + Y_w\sin\gamma - D_x \\ Y_c = \left(-X_w\sin\gamma + Y_w\cos\gamma\right)\cos\left(\alpha+\beta\right) \\ \qquad + Z_w\left(\cos\alpha\sin\gamma + \sin\alpha\cos\beta\right) \\ \qquad -\left(D_y\cos\alpha + D_z\sin\alpha\right) \\ Z_c = \left(X_w\sin\gamma - Y_w\cos\gamma\right)\sin\left(\alpha+\beta\right) \\ \qquad + Z_w\left(\cos\alpha\cos\beta - \sin\alpha\sin\gamma\right) \\ \qquad + D_y\sin\alpha - D_z\cos\alpha \end{cases} \tag{2}$$

where α is the angle between the axis O_cZ_c and the axis O_wZ_w. β is the angle that the axis O_cZ_c rotates around the axis O_cX_c (so the pitching angle is $\alpha+\beta$), γ is the angle that the axis O_cX_c rotates around the axis O_cZ_c. When the vehicle moves, β and γ can be obtained from the inertial navigation system(INS) on the ALV. (D_x, D_y, D_z) is the coordinate value of the origin of the C-CS in the W-CS. If (X_c, Y_c, Z_c), (D_x, D_y, D_z) and (α, β, γ) are known, (X_w,Y_w,Z_w) can be obtained, and a 3D map is generated.

In general, the characteristic of the scene and the requirements of the applications must be considered in representing 3D map. In autonomous navigation, we usually use grid-based representation [8],[1], which can effectively represent the scene and provide enough obstacle information for planning and navigation. However, the traditional method has several disadvantages in 3D map representation as follows.

(1)The distribution of the built 3D map is

inevitable to be sparse and non-uniform, because triangulation is a non-linear operation. This results in too much data in the closer grids and too few data or even none in the further grids. Therefore, many 3D points are computed repeatedly in closer grids, while precision in the further grids is reduced for less data. And even many further grids without reconstructed data have to be filled by interpolation.

(2) Traditional methods have to compute all pixels to yield dense disparity map. In fact, many pixels are in uninterested region and not contributive to map building, so finding their correspondences increase unnecessary computation.

(3) Mismatched pixels in disparity map make incorrigible errors in 3D map building.

To improve the defects in traditional methods, we propose a novel method based on PVHL.

3. The proposed method

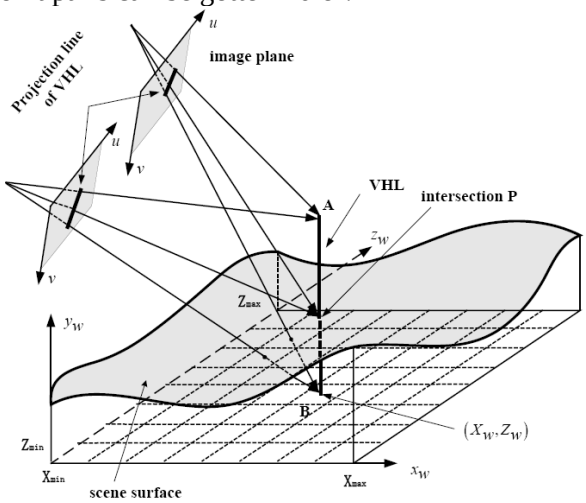
The stereo matching procedure is reverse to the 3D map building procedure. The ambiguity of stereo matching problems is difficult to be solved, which comes from the fact that the loss of information incurred when a 3D scene is projected on a 2D image. The method based on PVHL , which follows the imaging process and is similar to the hypothesis-test principle in mathematic statistics.

3.1 Principle of virtual height line(VHL)

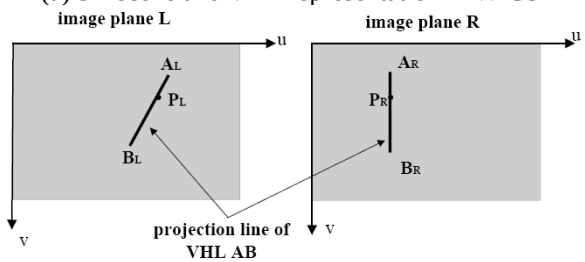
Once a stereo vision system is settled, its probably visible interested scene scope in the W-CS can be estimated, especially for ALV. In this method, we suppose that the interested scene scope for the ALV is the region delimited in $x_w z_w$ plane from Z_{min} to Z_{max} in the z_w direction and from X_{min} to X_{max} in the x_w direction. Then partition the region uniformly into grids with a certain size, as shown in Figure 2(a). We assume any scene grid (x_w, z_w) in W-CS has the same height. If the size of grids is small enough, the assumption is always valid. Without loss of generality, we use the height of the center point (X_w,Z_w) to represent that of the grid. Here, out of simpleness in analyzing the question, we introduce a VHL AB which has height but does not exist in the scene. The probable height range of AB is foreknown approximately from the inherent characteristic of the ALV. VHL AB is through the center point, and vertical to the $x_w z_w$ plane, and the intersection point of AB and the scene surface is indicated by P, then the height of the point P is also the height of grid (X_w,Z_w). Using equation (2) and equation (3), VHL AB is projected in left image and right image respectively, and two PLs $A_L B_L$ and $A_R B_R$ are generated, as shown in Figure 2(b). We suppose that projection of AB is based on left image and right image after epipolar rectification, and the projected points in stereo images have equal vertical coordinate. P is the only existent and visible point on AB, that is to say, only P_L on $A_L B_L$ and P_R on $A_R B_R$ are real correspondence pairs, and others are fake pairs. So the correspondence (P_L, P_R) has the maximum similarity metric, and others have relative smaller ones. Finding the height of P is equivalent to proving whether (P_L,P_R) has the maximum similarity metric or not, and this can be resolved by common stereo matching methods. When the height of each grid is obtained, the local 3D map is built.

$$\begin{cases} U = \dfrac{X_c d}{b} + U_0 \\ V = \dfrac{Y_c d}{b} + V_0 \\ d = \dfrac{bf}{Z_c} \end{cases} \tag{3}$$

We use the normalized ZNCC as similarity metric. Perhaps more than one pair points have bigger similarity metric for window effect. We must use some stronger constraints to improve the process of finding the best match point pair so that candidate point pairs can be gotten rid of.



(a) 3D scene and VHL representation in W-CS



(b) PLs of VHL AB in left and right image

Figure 2 Principle of VHL imaging in binocular stereo vision system

3.2 Constraints

(1)Maximum similarity metric constraint in horizontal maximum disparity range. The correspondence should have maximum similarity metric in all point pairs both in vertical direction on two PLs and in horizontal disparity range in *u* axis.

(2)Similarity metric continuity constraint. Because ZNCC is based on window, pixels surrounding the correspondence on the PLs have bigger similarity metric. But this is only a necessary condition not a sufficient condition.

(3)Height uniqueness constraint. No more than one correspondence exists on a PL. If no correspondence exists, the grid is in the occlusion region.

(4)Occlusion constraint. The correspondence can represent a 3D point only. If the position of a pixel projected from a further scene point and one from a closer scene point are identical, the far scene point is occluded by the closer. In other words, if an obstacle with a certain height is built, there must be an occlusion region behind the obstacle.

From the principle of the proposed method, we can see that partitioning the interested regions on $x_w z_w$ plane and limiting the height range of VHL, we can effectively avoid the computation of pixels in uninterested regions, such as sky. It not only reduces the unnecessary computation, but also gets rid of the incorrigible errors caused by mismatch pixels in uninterested regions.

4. Experiments

Figure 3 Left image of stereo vision system

To show how the proposed algorithm works, simulations and experiments are given to verify its performance. The ALV and the stereo vision system setting are shown in Figure 1. The image size in the experiments is 320*240. The first experiment for the scene is shown in Figure 3. The distances from the two obstacles to the camera are 700cm and 1000cm, and the height of them are 680mm and 750mm respectively. The built region is defined as follows: the range of Xc is from –450cm to 450cm, the range of Zc (depth) is from 600cm to 1500cm, and the range of Yc(height) is from –2000mm to 2000mm. And the size of each grid is 30cm*30cm.

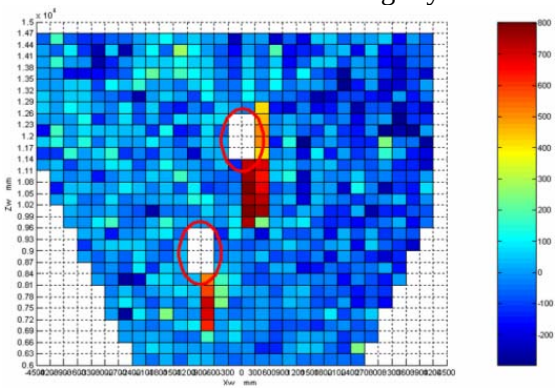
(a) PLs of VHLs in one row grids in the left image

(b) PLs of VHLs in one row grids in the right image

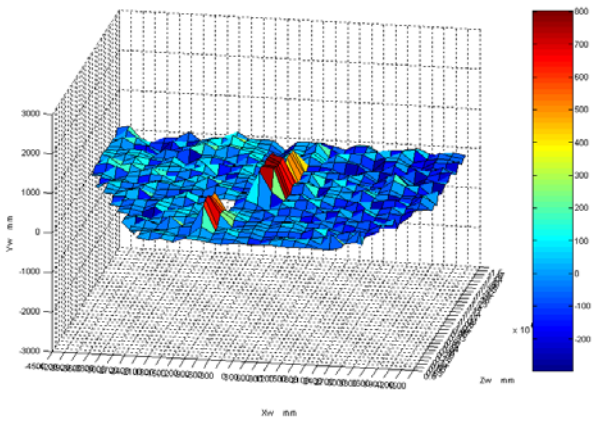
Figure 4 PLs of VHLs in one row grids

In our application, the proposed method is written in C++ language in Linux Red Hat 9, and the onboard computer is Intel Pentium 4 2.4GHz. The overall processing time is about 60ms, which satisfies the requirements of real-time automatic navigation. White lines in Figure 4 are the PLs of VHLs in one row grids. Most coordinates (*U*,*V*) from equation (3) are improved by sub-pixel interpolation because they are not integral. The result is shown in Figure 5. Two red circled regions as shown in Figure 5(a) are occlusion regions which may help to prove obstacles existence in closer vicinity. And Table 1 lists the comparison between the built data and the real data about the two

obstacles. The computation time, including stereo matching using ZNCC and the traditional map building, is more than 30s under condition where no speedup measures were acted on the process. Furthermore, we find that more than 90% computation time are consumed on stereo matching. And the result of the traditional is shown in Figure 6 and Table 2. Compared with the proposed, the performance of the traditional is slightly inferior.



(a) Platform for the local built 3D map



(b) the local built 3D map

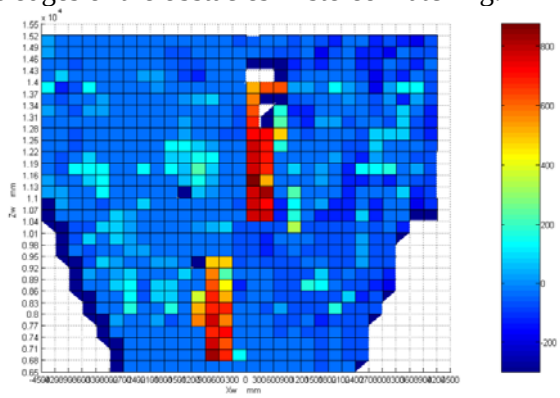
Figure 5 3D map using the proposed method for the scene as shown in Figure 3

Table 1. Comparison of results between the proposed and the real

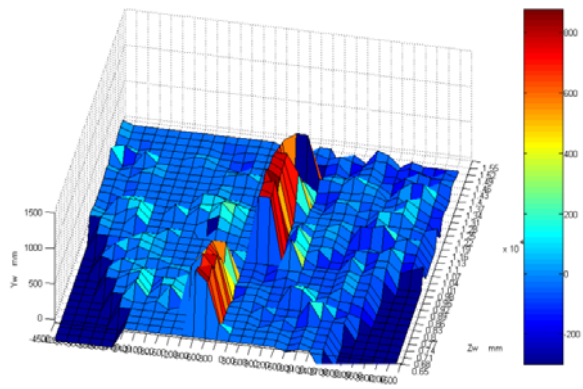
obstacle	real depth (cm)	built depth (cm)	real height (mm)	built height (mm)
1	700	690	680	624
2	1000	960	750	750

1-the closer obstacle, 2-the further obstacle

Another scene for the experiment is shown in Figure 7. The distances from the three obstacles to the camera are 600cm, 800cm and 1000cm, and the corresponding height are 590mm, 680mm and 750mm respectively. The results for the proposed and the traditional are shown in Figure 8 and Figure 9. And the built data are list in Table 3 and Table 4. The results show that the proposed has better performance than the traditional. In Table 4, the obvious errors in the height of the 3rd obstacle and at the distance of the 2nd obstacle result from the mismatches which come from the window effect at the edges of the obstacles in stereo matching.



(a) Platform for the local built 3D map



(b) the local built 3D map

Figure 6 3D map using the traditional method for the scene as shown in Figure 3

Experiment results show that the built 3D map reflects the actual scene basically and meets the need of ALV navigation. And from the experiments, we can see that stereo matching increases too much unnecessary computation time and that mismatches in the disparity map reduce the precision of the built 3D map. Besides, in the proposed method we suppose the built 3D position is known in advance, and then the corresponding pixel position is found in the image. So if the interested region is known, we can build this region only and no additional computation are wasted in uninterested regions. While the traditional is inevitable to compute all pixels to yield dense disparity map and build their positions.

Table 2. Comparison of results between the traditional and the real

obstacle	real depth (cm)	built depth (cm)	real height (mm)	built height (mm)
1	700	680	680	631
2	1000	1040	750	725

1-the closer obstacle, 2-the further obstacle

Figure 7 Left image of stereo vision system

5. Conclusion

In this paper, we analyze the defects of the traditional 3D map building method, and then propose a novel one based on PVHL. The proposed projects VHL into the stereo vision system by transforming the world coordinate of each point on VHL into its image coordinate, and generates two projection lines(PLs) in two images. On the two PLs, only the correspondence that is the projection of the intersection point of VHL and the surface of the scene has the maximum similarity metric. Then finding the height of the scene point is equivalent to proving whether the real correspondence of VHL has the maximum similarity metric or not. Simulation and experimental results are given to show that the built 3D map by the proposed is effective. And it is highly real-time and meets the need of ALV navigation. Compared with the traditional, the proposed has better performance in building 3D map, and it can build the interested region only, and reduce the unnecessary computation time.

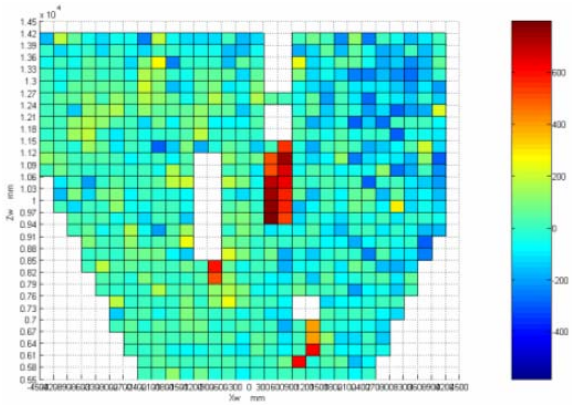
Acknowledgments

This work is supported by Natural Science Foundation of China under No.60702018.

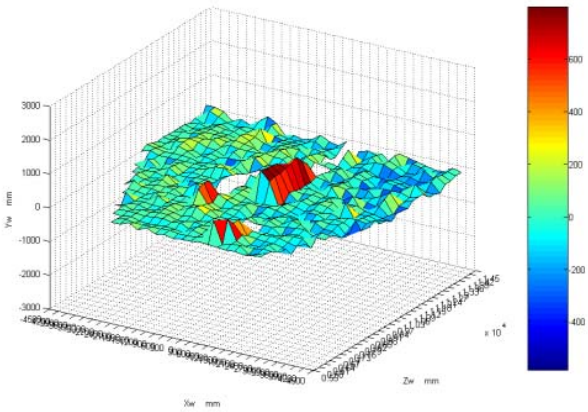
Table 3. Comparison of results between the proposed and the real

obstacle	real depth (cm)	built depth (cm)	real height (mm)	built height (mm)
1	600	580	590	612
2	800	790	680	587
3	1000	940	750	795

1-the closest obstacle, 2-the closer obstacle, 3-the furthest obstacle

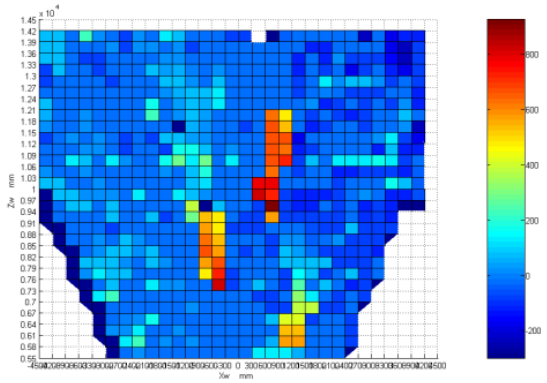


(a) Platform for the local built 3D map

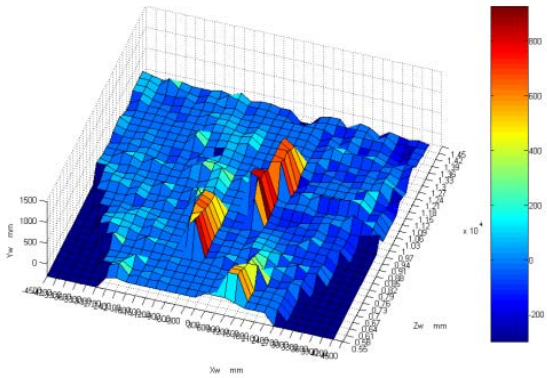


(b) the local built 3D map

Figure 8 3D map using the proposed method for the scene as shown in Figure 7



(a) Platform for the built local 3D map



(b) the built local 3D map

Figure 9 3D map using the traditional method for the scene as shown in Figure 7

Table 4. Comparison of results between the traditional and the real

obstacle	real depth (cm)	built depth (cm)	real height (mm)	built height (mm)
1	600	580	590	575
2	800	730	680	825
3	1000	910	750	575

1-the closest obstacle, 2-the closer obstacle, 3-the furthest obstacle

References

[1] H.H. Chen and Z.Z. Xu. "Local 3D Map Building and Error Analysis Based on Stereo Vision.," In :*Proceedings of The 31st Annual Conference of the IEEE Industrial Electronics Society*, pp. 379-382, 2005.

[2] J. M. Saez and F. Escolano,"A global 3D map-building approach using stereo vision," In: *Proceedings of the 2004 IEEE International Conference on Robotics and Automation*, pp.1197-1202, 2004.

[3] M.W. M. G. Dissanayake, P. Newman., S. Clark, et al.,"A solution to the simultaneous localization and map building (SLAM) problem," *IEEE Transactions on Robotics and Automation*,vol.17, no.3, pp.229-241,2001.

[4] J. H. Kim and S. Sukkarieh, "Airborne simultaneous localisation and map building," In: *Proceedings of the 2003 IEEE International Conference on Robotics and Automation*,pp.406-411,2003.

[5] M. Z. Brown, D. Burschka, and G. D. Hager. "Advances in computational stereo.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8): 993-1008, 2003.

[6] V. Kolmogorov, C. Rother. "Comparison of energy minimization algorithms for highly connected graphs," In: *European Conference on Computer Vision*, pp.1-15,2006.

[7] Q. Yang, L. Wang, R. Yang, et al. "Real-time Global Stereo Matching Using Hierarchical Belief Propagation," In: *The British Machine Vision Conference*, pp. 989-998,2006.

[8] I. S. Kweon. *Modeling Rugged Terrain by Mobile Robots with Multiple Sensors*. Carnegie Mellon University. Ph.D. dissertation, 1991.

[9] I. S. Kweon, and T. Kanade. High-Resolution Terrain Map from Multiple Sensor Data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp.278-292, 1992.