Ground Plane Stereo for Obstacle Detection

Kwang Hee Won and Soon Ki Jung
School of Computer Science and Engineering
Kyungpook National University
1370 Sankyuk-dong, Buk-gu, Daegu, 702-701, South Korea
Email: khwon@vr.knu.ac.kr, skjung@knu.ac.kr

***Abstract*—We propose a dense stereo matching algorithm for obstacle detection by using the ground plane assumption in which the scene includes objects standing on the ground plane. The algorithm produces a dense disparity map and its zero-valued pixels indicate the ground plane of the input image. The algorithm refines disparity map by filtering out the patterns of disparity values, which violate the ground plane assumption. Our algorithm can produce accurate disparity map for a given scene than the normal stereo algorithms with limited disparity budget and the disparity map is directly applicable to the obstacle detection task.**

I. Introduction

Stereo vision has become one of the most promising and potential sensors for developing intelligent vehicles due to the low cost and rich information that cameras provide. The scene needs to be analyzed properly and precisely to notify any event or to augment additional information on the driver's view for safe driving. In our case, we focus on the detection of obstacles like pedestrians or other vehicles in front of the vehicle.

In our paper, we assume that the vehicle is running on a normal road (a flat ground plane) and a stereo camera system is mounted on the vehicle. We also assume that the relative pose of the camera system with respect to the ground plane is given by the initial calibration procedure. Now, the obstacle detection tasks can be regarded as non-ground region search in the scene. With this problem statement, we define the ground plane by the homography from pixels on the projected ground plane in right image to the left image. In this concept, the hogmography applied pixels of the ground plane is well matched to the pixels on the other image. However, the pixels of an upstanding object have displacements which can be considered as disparity as shown in Fig. **??**.

Fig. 1. An overlayed left image and a ground plane homography applied right image.

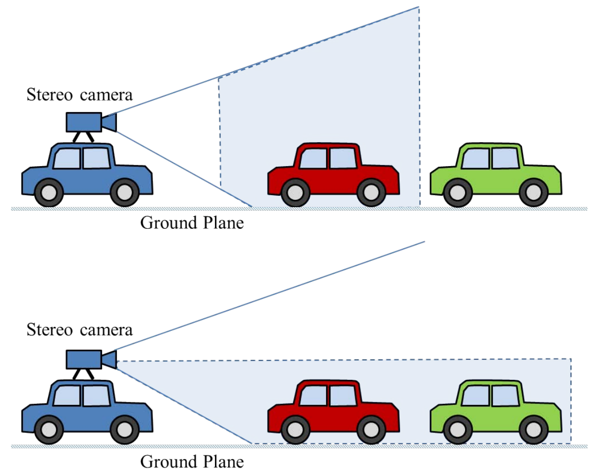


Fig. 2. Depth perception with limited disparity budget: (first row) conventional stereo matching algorithm, and (second row) ground plane stereo matching algorithm.

The typical stereo matching algorithms assume that the zero-disparity plane is parallel to one of the image planes. However, this assumption may not be effective under the observation that the vehicles and other objects necessarily stand or move around on a certain ground plane. The ground plane can be a good alternative as the zero-disparity plane for several reasons. First, the fixed disparity budget can be used more effectively because the dispairty values of the ground plane are the smallest possible values that each pixel can have. So, the longer distance can be represented with the fixed disparity budgets as shown in Fig. **??**. Next, many applications compute the height image to detect objects by accumulating the projected 3D points on the ground plane. However, the ground plane based disparity can be utilized directly to the object detection tasks. Lastly, the bad pixels can easily be removed by evaluating whether the pixel is a part of the ground-standing object or not.

In this paper, we propose a dense stereo matching algorithm based on the ground plane homography. The algorithm produces disparity maps which have zero-disparity values for the ground plane. The proposed algorithm also effectively filters out the matching failures using the ground plane scene assumption.

This paper is organized as follows. In Section **??**, we introduce the related researches about the stereo algorithms that the homography is applied. In Section **??**, we describe the ground plane based stereo matching algorithm. In Section **??**, we show some experimental results, and in Section **??**, we conclude this paper.

II. Related Works

The literature concerning about object detection and unmanned automatic vehicle researches the utilization of ground plane homography.

Arróspide et al. made use of ground plane homography of single-view consecutive images to detect moving objects on the road [**?**]. They obtained the homography by feature matching with prior knowledge. Se and Brady extracted edges from stereo images and matched them to detect objects on the ground [**?**]. They employed the Kalman filter to trace the edges, and estimated the ground plane using RANSAC. Similarly, Xie extracted objects by classifying the overlapped contours obtained by applying ground plane homography to one of stereo images [**?**]. We try to find ground plane based dense disparity map instead of feature correspondences. Our algorithm employs the Semi-Global Matching algorithm [**?**] to aggregate matching costs and finds globally optimal disparity values along the multiple directions. The dense disparity map is useful because some patterns (such as the increasing disparity from the ground plane) can be applicable to detect objects.

Williamson and Thorpe addressed the use of multiple homography matrices to calibrate multiple camera to obtain more robust matching costs from the warped template according to the homography [**?**]. They introduced the use of homography of both parallel and perpendicular to the camera axis, and combined one. Similarly, Williamson and Thorpe compared the matching costs of horizontal and vertical stereo matching costs and extracted objects [**?**]. They made groups of pixels that have the similar depth values and location, and filtered out a small sized groups. They used the multicamera system with multiple homography to enhance the quality of the results, however, our algorithm performs the aggregation process with the costs from ground plane homography. Besides, our algorithm discards bad pixels by validating whether the pixel is a part of standing objects or not. After that, we detect each connected component on the disparity map as objects.

III. Ground Plane Stereo Matching

Fig. **??** shows the relationship between the stereo camera system installed on a vehicle and the ground plane Π_G. Π_L and Π_R represent the projected plane of Π_G on each image plane, respectively. Then the relation between Π_L and Π_R, with respect to the Π_G can be denoted by a single homography, H_{RGL} and expressed as,

$$H_{RGL} = H_{GL} \times H_{RG}, \tag{1}$$

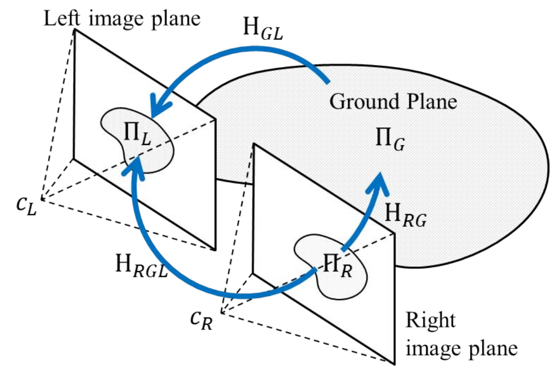


Fig. 3. Ground plane homography.

where H_{RG} and H_{GL} are the homographies from Π_R to Π_G, and from Π_G to Π_L, respectively. Fig. **??** shows an overlapped example of a left image and its corresponding H_{RGL} applied right image.

In this ground plane stereo geometry, the matching cost of a pixel p with disparity value d is represented as follows,

$$C(p, d) = |I_L(p) - I_R(H_{RGL}(p) + d)|, \tag{2}$$

where I_L and I_R are the intensity values of the given pixel. The value d in equation **??** represents the ground plane disparity. By adding $H_{RGL}(p)$ to d, the ground plane based disparity values can be converted to the typical disparity values. $I_R(H_{RGL}(p)+d)$ is obtained by interpolation because $H_{RGL}(p)$ often indicates the sub-pixel position. The better (more discriminative) matching costs can be obtained from the homography based intepolation for the ground plane, a horizontal surface [**?**], [**?**].

The matching costs are computed for all disparity ranges and aggregated by the equation from the Semi-Global Matching algorithm (in short, SGM) [**?**]. SGM aggregates matching cost along various directions instead of 2D global optimizations, using the following equation [**?**], [**?**],

$$\begin{aligned} L_r(p, d) = C(p, d) & \\ + \min\Big(& L_r(p_{-1}, d), L_r(p_{-1}, d \pm 1) + P_1, \\ & \min_i L_r(p_{-1}, i) + P_2\Big), \end{aligned} \tag{3}$$

where $L_r(p, d)$ is an aggregated matching cost of pixel p and disparity d. p_{-1} represents the precedence pixel along the direction of aggregation. P_1 and P_2 are the penalty for the small (± 1) and large $(1 <)$ disparity changes. The P_2 is determined adaptively according to the difference of pixel values from P_1 to P_2.

The optimal disparity is then determined as the disparity value that has the smallest sum of aggregated costs for all directions. After the initial disparity map is computed, the median filter is applied to remove the flicking noise. Next, we filter out the patterns of disparity values that do not fit the ground plane assumption. In outdoor environments, artifacts can be produced for various reasons such as the different exposure, motion blur (especially the closer part of the scene), the occlusion, asynchronous acquisition and so on.

The filtering rules are simple: 1) the pixels which have the disparity values less than ϵ_g, the small value, are considered as ground pixels, and 2) if the difference of the disparity values between two adjacent pixels are less than ϵ_c, another small value, those two pixels are considered to be connected each other. Lastly, 3) any pixels which do not have the connected path to any ground pixels using the rule 1) and 2) are filtered out. The rules can be represented by the recursive definition of $\delta(p)$ function which has a value of 1 for accepted pixels and 0 for rejected pixels. The $\delta(p)$ can be represented as follows,

$$\delta(p) = \begin{cases} \text{if } D(p) < \epsilon_g, & 1, \\ \text{if } D(p) \geq \epsilon_g \text{ and } N_p = \phi, & 0, \\ \text{otherwise} & \max_i\{\delta(q_i \in N_p)\}, \end{cases} \tag{4}$$

where $D(p)$ is the ground plane based disparity of pixel p and N_p is a set of non-visited neighbor pixels within radius r and which have the acceptable difference of disparity value ϵ_c with the disparity value of pixel p.

We show the refinement result using the filtering rules in Fig. **??**. The equation (**??**) is represented in a recursive form, but we implemented it using multiple scans and merge for the computational efficiency. The obtained disparity map can be transformed to the traditional disparity map by inversely applying the homography relation in equation (**??**). Otherwise,

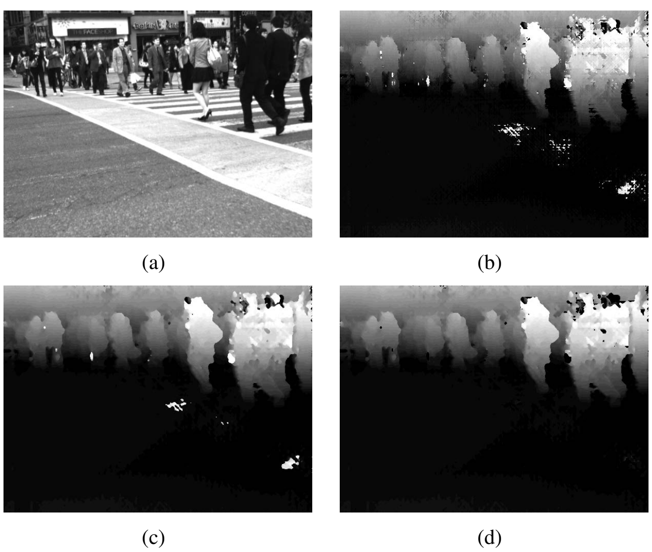


Fig. 4. Disparity map refinement: bad pixels that violate the ground plane constraint are removed. (a) Left image. (b) Initial disparity map. (c) Median filter (5×5) applied. (d) Equation (**??**) applied.

it can be used directly by finding components using the discontinuity of disparity values on the object boundary.

IV. Experimental Results

We used a stereo camera system which includes two CCD cameras, with 6 mm C-mount lens on a 200 mm-horizontal rig. Then, we obtained stereo video from a moving vehicle

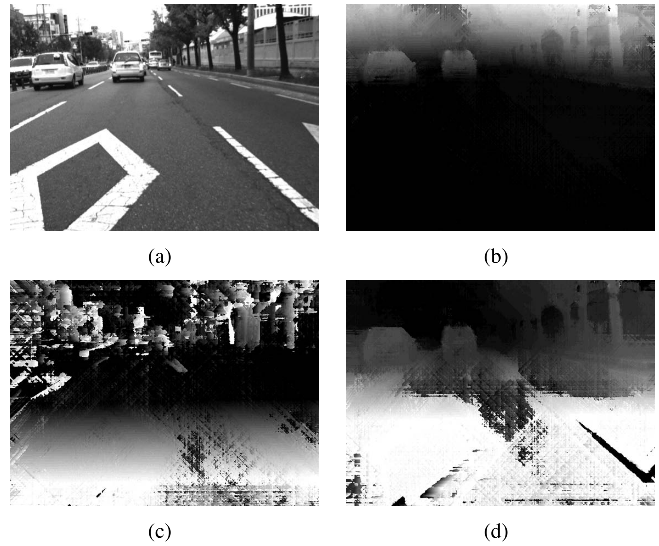


Fig. 5. Ground plane stereo matching: **??** is the input image. **??** is the ground plane based disparity map which is obtained using 32-level disparity values. **??** and **??** show the 32-level disparity range cannot represents the whole scene with the traditional disparity map.

and computed its disparity maps. The resolution of the input is 640-width and 480-height and the 20-pixel boundaries of each side are discarded.

We first compared the visual quality of the disparity maps which are obtained by the ground plane based stereo and typical stereo matching. We implemented both the proposed algorithm and one of the typical stereo matching algorithm which produces the perpendicular plane based disparity map. The homography applied matching cost is used for the proposed algorithm. The difference of grayscale pixel values is used as the matching cost, and the SGM type cost aggregation is used for both implementation. The left-right consistency check is omitted in both implementations because it can be applied in both algorithms. The penalty parameters P_1 and P_2 are adjusted experimentally to produce the best result. The homography of the ground plane is obtained by assigning small number of pixel correspondences manually for the initial frame. For the other frames, (as we assumed in Section **??**) the geometric relation between the vehicle with stereo system and the ground plane is the same. The automated acquisition of the precise ground plane homography is also an important issue, and many existing approaches can be applyed.

Fig. **??** shows an example of disparity map resulted from the ground plane based stereo matching and the typical stereo matching. Fig. **??** is the ground plane based disparity map obtained by the proposed algorithm. The darker represents the smaller values, and the brighter represents the higher values. The ground plane in this scene has the small values and the upstanding objects have the increasing disparity values starting from the bottom to the top. Fig. **??** and **??** shows the results of the typical vertical plane based stereo matching algorithm. The smaller value can be interpreted as the farther distance from the camera for the typical disparity maps. To

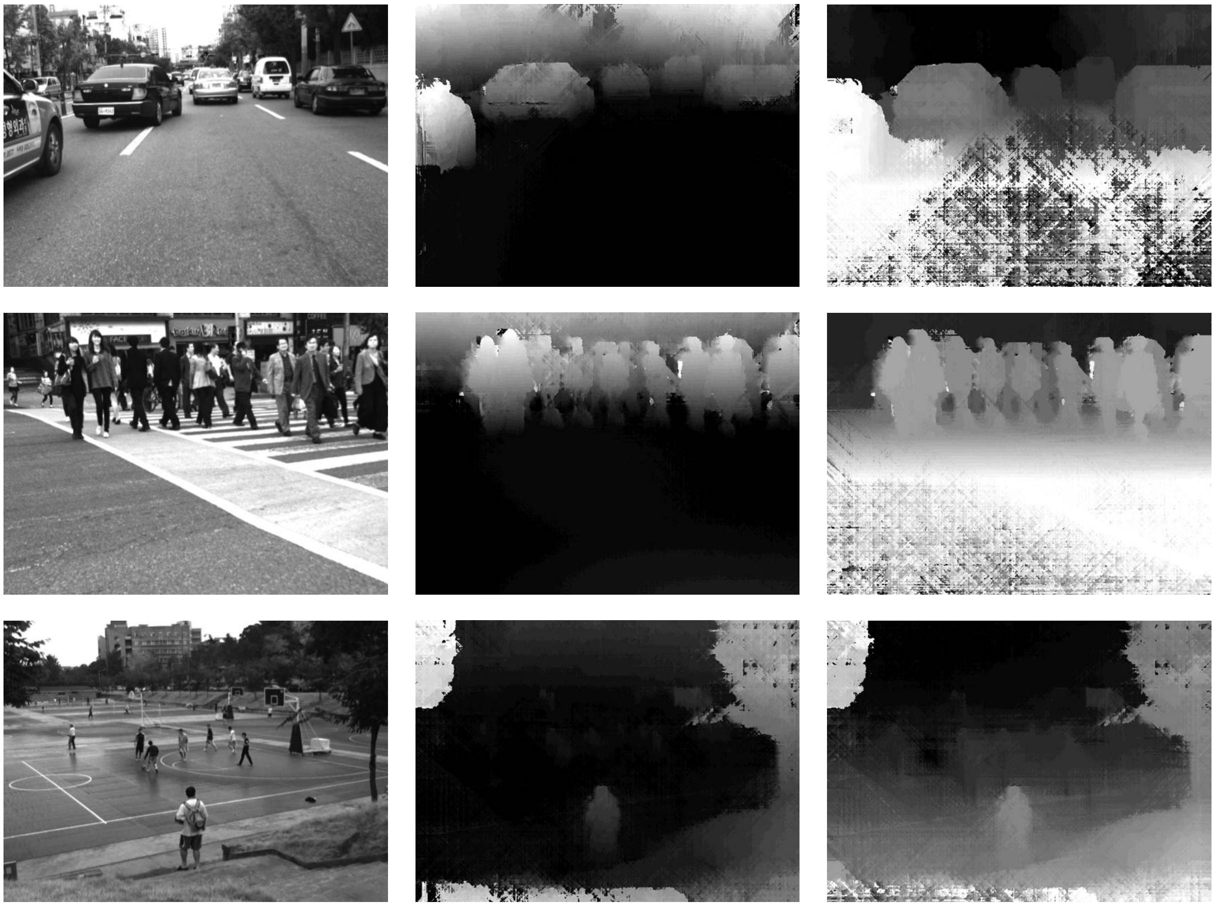


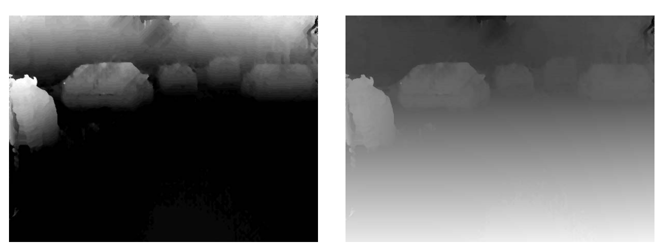
Fig. 6. Ground plane scene with various objects and the corresponding disparity maps. The first column shows the input left image. The second column shows the result of ground plane stereo matching. The last column shows the result of SGM with 32-level disparity.

obtain Fig. **??** and **??**, the 32-level of disparity range are applied for the close and distant objects, respectively. The matching failures are shown because the required disparity budgets are larger than the given disparity range for typical disparity maps. However, the proposed algorithm produces the clearer ground regions with objects which are distinguishable from the ground and other background objects. In case of object regions, the proposed algorithm tend to produce less clear results than that of the typical algorithm because the proposed algorithm uses the homography applied pixel values to compute matching costs. The homography applied pixel values degrade the discriminative power of the matching costs for the vertically facing objects with cameras.

Fig. **??** shows more results with various ground plane scenes with vehicles, pedestrian, and other obstacles. The same levels of disparity range with Fig. **??** are used. For the results of the third column, the 32-levels of disparity values are adjusted to detect the middle and far objects. The same homography with the initial frame is used for the first and second row because the scene is captured from the same vehicle, and the different homography matrix is computed for the third row.

The ground plane based disparity map can be easily transformed to the typical disparity values by adding the disparity values of the ground plane which can be obtained using the homography. Fig. **??** shows an example of the ground plane based disparity map and its transformed one. In Fig. **??**, the transformed disparity map contains over than 90-levels of disparity values. It is almost three times larger ranges compare to the 32-levels disparity which have used to obtain Fig. **??**.

The ground objects, which exist on the ground plane, can be extracted by removing the ground plane and the distant objects and background. We remove the ground area using the ground plane based disparity values using the same threshold value, ϵ_g in equation **??**. Next, we remove the distant objects which are

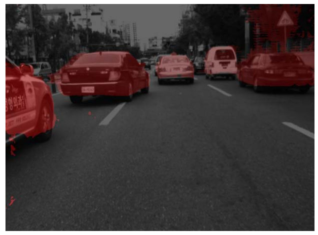


(a) (b)

Fig. 7. Ground plane based dispariy map and transformed disparity map. (a) ground plane disparity map. (b) transformed disparity map.

not in the interested range, using the transformed disparity values like in Fig. ??. The example of extracted objects areas and their disparity values is represented in Fig. ??. The gray regions represent the transformed disparity values

(a)



(b)

Fig. 8. Object detection using ground plane disparity map. (a) detected object regions. (b) overlayed objects on the left image.

and the values can be translated into the distance using the stereo geometry. We superimposed the object areas on the input image as shown in Fig. ??. The result shows that the ground plane based disparity map can provide clues about the areas of ground plane and the distances of ground objects. The computation time of the proposed algorithm is a few seconds and the realtime performance can be accomplished with parallel processing by utilizing the multi-core CPU or GPU [?]. Moreover, the proposed algorithm requires one third of disparity budget than the typical stereo matching algorithm and the computation time increases according to the increasing disparity levels.

V. Conclusion

In this paper, we proposed a dense stereo matching algorithm to produce ground plane based disparity values. We also proposed a filtering method for the abnormal disparity patterns in the ground plane scene. The proposed algorithm produces more abundant disparity on typical ground vehicle scenes and the disparity values can directly be used for the object detection task.

However, the moving of vehicles toward the rapid incline will violate the ground plane assumption. In this case, the relative pose of the camera to the ground plane need to be updated properly. We will develope the algorithm that refines the homography according to the rapid changes of the slope of the ground plane using the ground plane disparity maps.

That the ground plane based disparity values provide good estimations of the ground plane, but it produces rather weak clues for the vertical surfaces than the typical stereo algorithms. In the future, we will combine the strong points of both methods to produce better disparity maps, which facilitates as important evidence for object detection.

Acknowledgment

This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the CITRC(Convergence Information Technology Research Center) support program (NIPA-2011-C6150-1102-0011) supervised by the NIPA(National IT Industry Promotion).

References

[1] J. Arróspide, L. Salgado, M. Nieto, and R. Mohedano, "Homography-based ground plane detection using a single on-board camera," *IET Intell. Syst.*, vol. 4, no. 2, pp. 149–160, 2010.

[2] S. Se and M. Brady, "Stereo vision-based obstacle detection for partially sighted people," in *Proc. Third Asian Conference on Computer Vision*, 1998.

[3] M. Xie, "Matching free stereovision for detecting obstacles on a ground plane," *Machine Vision and Applications*, vol. 9, pp. 9–13, 1996.

[4] H. Hirschmüller, "Accurate and efficient stereo processing by semi-global matching and mutual information," in *Proc. CVRP*, 2005, pp. 807–814.

[5] T. Williamson and C. Thorpe, "A specialized multibaseline stereo technique for obstacle detection," in *Proc. CVPR*, 1998.

[6] T. Williamson and C. Thorpe, "Detection of small obstacles at long range using multibaseline stereo," in *Proc. of the IEEE International Conference on Intelligent Vehicles*, 1998.

[7] V. Kolmogorov and R. Zabih, "Computing visual correspondence with occlusions using graph cuts," in *Proc. ICCV*, vol. 2, 2001, pp. 508–515.

[8] J. Sun, Y. Li, S. Bing, and K. H. yeung Shum, "Symmetric stereo matching for occlusion handling," in *Proc. CVPR*, 2005, pp. 399–406.

[9] K. H. Won and S. K. Jung, "*h*SGM: Hierarchical pyramid based stereo matching algorithm," in *Proc. Advances Concepts for Intelligent Vision Systems*, 2011.