

Integrating LIDAR into Stereo for Fast and Improved Disparity Computation

Hernán Badino, Daniel Huber and Takeo Kanade
Robotics Institute
Carnegie Mellon University
Pittsburgh, USA

Abstract—The fusion of stereo and laser range finders (LIDARs) has been proposed as a method to compensate for each individual sensor's deficiencies – stereo output is dense, but noisy for large distances, while LIDAR is more accurate, but sparse. However, stereo usually performs poorly on textureless areas and on scenes containing repetitive structures, and the subsequent fusion with LIDAR leads to a degraded estimation of the 3D structure. In this paper, we propose to integrate LIDAR data directly into the stereo algorithm to reduce false positives while increasing the density of the resulting disparity image on textureless regions. We demonstrate with extensive experimental results with real data that the disparity estimation is substantially improved while speeding up the stereo computation by as much as a factor of five.

Keywords-stereo; LIDAR; sensor fusion; dynamic programming; disparity space reduction; path promotion; range images

I. INTRODUCTION

The fusion of stereo and laser range finders (LIDARs) has been proposed as a method to compensate individual sensor's deficiencies – stereo output is dense but noisy for large distances while LIDAR is accurate but sparse. By fusing the output of both sensors, accurate and dense range information can be obtained (Fig.1). Stereo usually performs poorly on textureless areas and on scenes containing repetitive structures. These situations reduce the density of the output and can lead to the generation of surfaces that actually don't exist. In this paper, we propose the early integration of LIDAR data *into* the stereo algorithm by providing an a priori estimate of the disparity.

We propose two novel ways of LIDAR/stereo integration: *disparity space reduction* and *path promotion*:

Disparity Space Reduction. The space of possible stereo disparities is reduced by defining an appropriate disparity interval for *each* pixel in the image based on the more precise LIDAR range estimates. To accomplish this, the sparse LIDAR data are transformed into a dense range image. The range image is used to predict maximum and minimum disparity images, taking into account the noise properties of the LIDAR measurements. Our experiments show that this process results not only in more precise disparity images, but also reduces the overall computational requirements, thereby speeding up the stereo computation

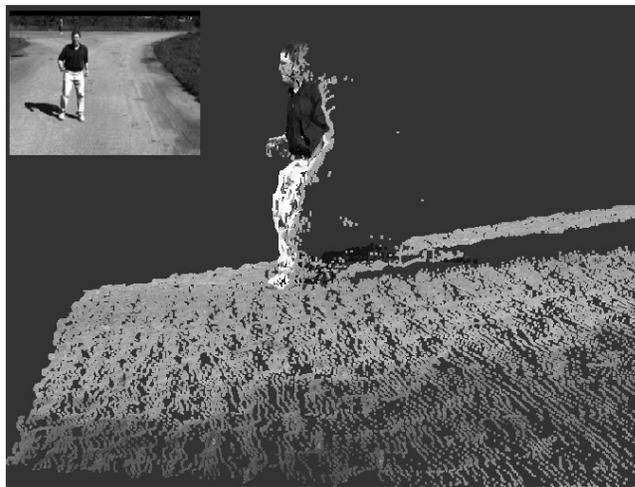


Figure 1. Result obtained with LIDAR/stereo fusion.

significantly.

Path Promotion. LIDAR range data is used to predict the expected dense disparity image and its gradient, both of which can be used as a prior for computing the final disparity. When using dynamic programming as the stereo computation method, predefined paths obtained from the prior can be promoted within the disparity search space resulting in fewer stereo artifacts.

A. Related Research

Multi sensor fusion has several advantages as noted in [1], and it has become standard nowadays in autonomous robotics systems. The general fusion of LIDARs with photogrammetry is addressed in [2], where the authors identify and review various levels of integration. However, the fusion of stereo with LIDARs or time-of-flight (TOF) cameras has only been approached a few times in the literature. There are mainly two types of stereo/TOF integration, a posteriori and a priori integration. A posteriori fusion aims at integrating the outputs (i.e., disparity maps) of both sensors to enhance the final output of the system [3], [4], [5], [6]. This is usually accomplished by using evidence grids [6] or at the object level [4], [3], [2]. A priori integration, on the other hand, fuses the ranges obtained from the TOF camera into the stereo computation algorithm in order to

improve the resulting disparity map. Bayesian approaches use the output of the TOF camera to provide a prior of the disparity image [7], [8], [9], [10]. Obtaining the maximum a posteriori estimate is time consuming and not suitable for real-time applications. [11] uses a volumetric approach, by which TOF cameras provide the initial estimate of the 3D structure, which is later refined by using multiple cameras. The method is accurate, but computationally expensive. In an analogous way, TOF cameras were used to provide an initial disparity estimate in a hierarchical approach in [12], to adapt window size based on previous range segmentation in [13], and to modify matching costs in a plane-sweeping algorithm [14].

In contrast to the above a priori fusion methods, which use exclusively TOF cameras in indoor environments, we propose the tight fusion of LIDAR and stereo to support real-time processing, by reducing the disparity search space and promoting paths with dynamic programming.

II. PRELIMINARIES

In this section, we introduce the basic concepts used throughout this paper.

Stereo Computation. We follow the four main steps identified in [15] to produce disparity images: 1) *Matching cost computation* – compute the similarities between the left and right images; 2) *Aggregation* – aggregate the resulting similarity values to reduce noise; 3) *Optimization* – find the “optimal” disparity image. Different optimality criteria lead to the definition of different optimization algorithms; 4) *Refinement* – reduce outliers and improve the disparity accuracy.

Depending on the similarity measure used to compute the disparity space image, the second step might be actually merged with the first or not performed at all.

Disparity Space Image. Matching cost computation requires computing the differences between pixels of the left and right views to produce a disparity space image (Fig. 3d). The disparity space image (DSI) is a 3D array where each component contains a similarity/dissimilarity measure between corresponding points in the left and right images. Specifically $DSI(u, v, d)$ is a scalar containing the similarity/dissimilarity between pixel (u, v) in the left image and $(u + d, v)$ in the right image (assuming that the images have been rectified into the standard left-right stereo configuration). The computational complexity of generating the DSI is linear with image size and disparity interval. It is critical to choose the disparity interval D in order to cover the expected distances to all objects in the scene. A large disparity interval ensures the observability of more ranges, but increases the computational requirements.

Spherical Range Image. We use range images to generate disparity predictions from LIDAR data. A spherical range image (SRI) is a function $s(\theta, \phi)$ in which the domain

(θ, ϕ) represents the azimuth and elevation components, and the codomain $s()$ is a real value that represents a distance or range (Fig. 3a, bottom). The three components of the image – azimuth (column), elevation (row) and range (image value) – define the coordinate vector of a 3D point, i.e., $(\theta, \phi, r)^T$ with $r = s(\theta, \phi)$.

Calibration. We assume the sensors have been calibrated to enable the appropriate geometrical transformations of LIDAR and stereo measurements.

III. DISPARITY SEARCH SPACE REDUCTION

Typical stereo algorithms define a single disparity interval D for the entire image. We propose to use the LIDAR data to predict an expected range interval for each pixel of the image. This idea has two benefits. First, it reduces errors in the computation of stereo because the optimization algorithm is less likely to get stuck in a local minima. Second, it reduces the computation time, which reduces linearly with the reduction of D .

The main difficulty is to efficiently transform sparse LIDAR range measurements into dense disparity intervals. We propose the following four step procedure (Fig. 3):

- 1) Calculate the spherical range image.
- 2) Apply maximum and minimum filters to the SRI.
- 3) Compute minimum and maximum disparity images.
- 4) Compute the reduced DSI.

We now describe these steps in more detail.

Step 1: SRI Calculation. SRIs are generated using the measurements $m_i = (\theta, \phi, r)^T$ from a LIDAR sensor. The range image position corresponding to coordinate (θ, ϕ) is identified, and the range r is stored at that position of the image. To consider the measurement noise, and the uncertainty produced by possible inaccuracies of time synchronization, calibration, and registration errors given by the motion of the scene while the data was acquired, we assume that the measurements are independent and affected by zero mean noise with variances $\sigma_\theta^2, \sigma_\phi^2, \sigma_r^2$.

Since the LIDAR measurements are usually not equally distributed over the measurement space, the resulting range image obtained is sparse (Fig. 2a). Linear interpolation is performed in two steps, first horizontally (Fig. 2b), then vertically (Fig. 2c). Holes are not filled across large gaps or depth discontinuities. A depth discontinuity exists between two range measurements r_i and r_j if the ratio of ranges, $\max(r_i, r_j)/\min(r_i, r_j)$, is larger than an empirically determined constant K (1.1 in our experiments). Observe from the interpolated SRI in Fig. 2c that depth discontinuities around the pedestrian (black regions) are correctly preserved.

Step 2: Minimum and Maximum Filters. The next step is to define the minimum and maximum possible range for each angular direction considering the noise properties of

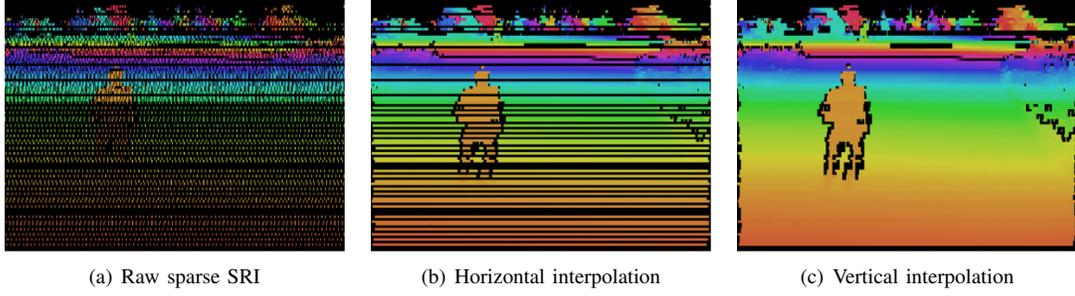


Figure 2. Vertical and horizontal interpolation. The color encodes the range.

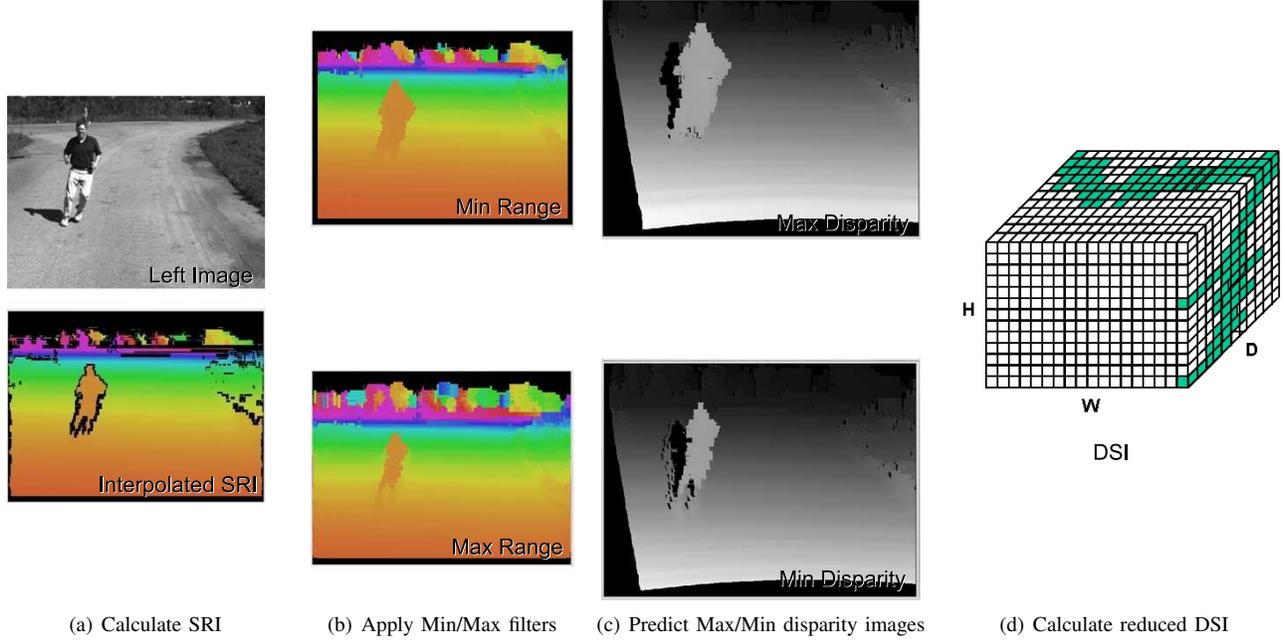


Figure 3. The four steps to generate a reduced Disparity Space Image. The color of the spherical range image encodes the range.

the measurements (Fig. 3b). The minimum and maximum filters are implemented using a two step procedure. First, for the min filter, an erosion morphological operation is applied using a rectangular-shaped structuring element. The analogous dilation operator is used for the max filter. The size of the rectangle should be chosen taking into account the noise properties of the measurements. We use a filter size of three times the standard deviation ($3\sigma_\theta \times 3\sigma_\phi$). Second, a scalar of size $3\sigma_r$ is added/subtracted to the resulting max/min SRIs to account for range noise.

Step 3: Prediction of Max/Min Disparity Images. Maximum and minimum disparity images are obtained from the minimum and maximum range images, respectively, providing an expected disparity interval for each pixel of the image (Fig. 3c). The LADAR-to-disparity transformation is performed by a remapping operation. Each pixel in the SRI, which corresponds to a 3D point (X, Y, Z) , is mapped to the image coordinate of its corresponding projection onto the left image. This is performed by using a look-up

table containing the corresponding image displacements. Ranges are then transformed into disparities by the equation $d = Bf/Z$, where d is the disparity corresponding to the depth Z , and Bf are the stereo system baseline and focal length.

Step 4: Reduced DSI Computation. Finally, a reduced DSI is computed by using an individual disparity interval for each pixel. The next three steps of the stereo computation (Section II) can now be applied to obtain the final disparity output using the reduced disparity space. Missing range data and occlusions can generate empty regions in the max/min disparity images (black regions on Fig. 3c). The full disparity interval D is used on those regions.

The above procedure reduces the overall computational requirements in the calculation of the DSI. Observe that the computational overhead of the first three steps is quite low, since they involve only simple filters and remapping operations that can be highly optimized in current CPUs and GPUs.

IV. PROMOTING PATHS IN DYNAMIC PROGRAMMING

The information obtained from LIDAR sensors can be used for more than just restricting the valid disparity interval. In this section we show how a simple dynamic programming (DP) optimization algorithm (Fig. 4) can be easily adapted to fuse image and LIDAR to substantially improve the final disparity image.

A. Dynamic Programming

Dynamic programming has been widely used for the fast computation of stereo [15], since it offers a good trade-off between efficiency and computation time. This section introduces the basic DP algorithm, which is adapted in the next section to include LIDAR predictions.

The energy function minimized by DP can be expressed in recursive form and usually contains a data and a smoothness term:

$$E(i, j) = \arg \min_k (E(i-1, k) + S(j, k)) + D(i, j) \quad (1)$$

where $E(i-1, k)$ corresponds to the energy function of a preceding node, $D(i, j)$ is the cost added to the path if the solution goes through the node (i, j) and $S(j, k)$ is a term penalizing the solution going from node (i, j) to node $(i-1, k)$.

In our basic implementation of the DP algorithm, we have chosen the following data and smoothness terms:

$$D(i, j) = DSI(i, v, j) \quad (2)$$

$$S(j, k) = \begin{cases} C_S \cdot |j - k| & ; \text{ if } |j - k| < t_S \\ C_S \cdot t_S & ; \text{ otherwise} \end{cases} \quad (3)$$

The data term is the corresponding value of the DSI for column i and disparity j for the current scan line v being evaluated. The smoothness term linearly penalizes the jumps in disparity up to a threshold. The threshold has the effect of assigning the same cost to disparity jumps larger than t_S , thus preserving depth discontinuities. The real potential of DP lies in the smoothness term. Without the smoothness term (i.e., $C_S = 0$), DP equals winner-take-all (WTA) optimization.

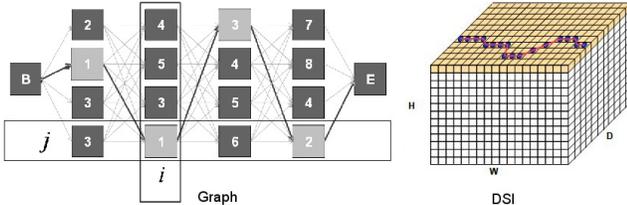


Figure 4. DP finds the optimal path in a directed graph without cycles. Columns of the graph represent image columns while rows represent disparities. Values in the nodes represent costs obtained from the DSI. DP is used to optimize the disparity computation at each row of the image.

B. Disparity Promotion

The ranges obtained from LIDAR can be used to compute expected disparities. The disparity prediction can help to find a solution in the DSI. The energy function described above can be expanded to include an additional term P penalizing the deviation of the solution from the expected value, i.e.,

$$E(i, j) = \arg \min_k (E(i-1, k) + S(j, k)) + D(i, j) + P(j, d(i)) \quad (4)$$

The last term is a penalty imposed on the energy function when the solution over node j deviates from the expected solution $d(i)$ for image column i . P has the same form as the smoothness term

$$P(j, d) = \begin{cases} C_P \cdot |j - d| & ; \text{ if } |j - d| < t_P \\ C_P \cdot t_P & ; \text{ otherwise} \end{cases} \quad (5)$$

The parameter C_P controls the confidence of the prediction obtained from LIDAR. If $C_P = 0$, no penalty is imposed, and the solution equals standard DP. When C_P is large, the solution equals the LIDAR prediction. The threshold t_P can be used to truncate the penalty imposed.

C. Disparity Gradient Promotion

The same concept of disparity (magnitude) promotion can be applied to gradients. The smoothness term S defined in Eq. 3 is motivated by the observation that the world is usually continuous, i.e., we assume that disparities vary smoothly almost everywhere. The penalty is applied to the change in disparity between consecutive columns (i.e., $|j-k|$ in Eq. 3), which is reasonable if the expected change in disparity is *unknown*. But if the disparity change is known or can be predicted, the penalty should be imposed only if disparity change is different from its prediction. Therefore, we propose to replace the smoothness term S with the following:

$$S(i, j, k) = \begin{cases} C_S \cdot |j - k - d'_u(i)| & ; \text{ if } |j - k - d'_u(i)| < t_S \\ C_S \cdot t_S & ; \text{ otherwise} \end{cases} \quad (6)$$

where $d'_u(i)$ is the expected disparity gradient at column i . The smoothness term has a minimal impact in the computational requirements and does not add any additional parameter to the energy function. In the case of missing



Figure 5. A custom stereo platform and a Velodyne LIDAR.

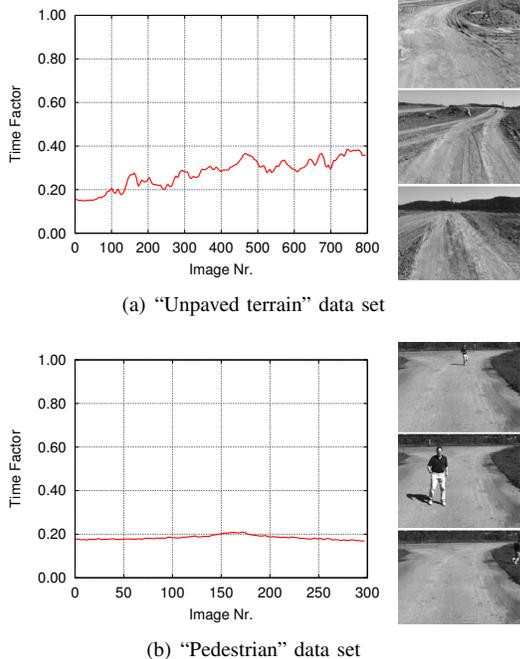


Figure 6. Relative time required by the reduced DSI algorithm.

information (i.e., no prediction), $d'_u(i)$ can be set to zero, equaling the original smoothness term.

D. Disparity and Gradient Prediction

To promote paths in the DSI space using the predictions of disparity magnitude, $d(i)$ and gradient $d'_u(i)$ must be first obtained. The estimation of a disparity image based on LIDAR can be obtained following the same procedure of Section III, but where the original SRI is used instead of a min/max SRI. The estimation for the gradient is obtained by applying a Prewitt filter on the resulting predicted disparity image (Fig. 7). The computation time of this step is negligible compared to the total time required for the computation of the final disparity image.

V. EXPERIMENTAL RESULTS

We validated the proposed algorithms using a sensor suite consisting of a Velodyne HDL-64E LIDAR and a custom-built stereo camera rig (Fig. 5). The Velodyne rotates at 10 Hz and records one million points per second using 64 horizontally scanned beams. The stereo rig uses two Flea2 Point Grey cameras operating at 10 Hz and 800×600 pixels resolution. The baseline of the stereo cameras is 33 cm with a focal length of approximately 1280 pixels. The sensors were mounted on the roof of an experimental test vehicle, and the data for the experiments was acquired while the vehicle was driving. The DP parameters C_S , t_S , C_P , and t_P of Eqs. 1-6 were empirically determined, while the variances σ_θ^2 , σ_ϕ^2 , and σ_r^2 for the max/min filters (Section III) were obtained from the LIDAR specifications.

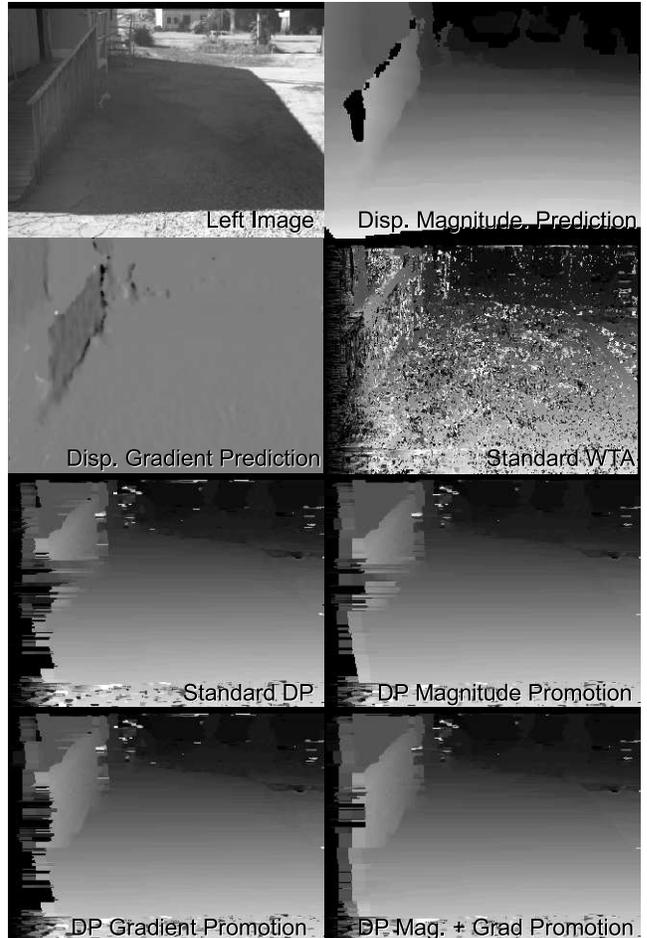


Figure 7. Benefit of path promotion on the DP algorithm. The gradient image is encoded from black (-1 disparity/pixel) to white (1 disparity/pixel).

A. Disparity Search Space Reduction

The use of LIDAR data to restrict the search of the disparity space allows not only a faster stereo computation, but also an improvement in the precision of the resulting disparity image. The next two subsections demonstrate this.

1) *Time Reduction*: We illustrate the speed improvements achieved with the proposed method using two different real sequences. As a baseline, we measured the computation time required for each stereo pair in each sequence for a fixed disparity range of 110 pixels. We then computed the time required by our disparity space reduction algorithm for the same sequences. The relative time for the proposed method with respect to the baseline is shown in Fig. 6. Depending on the depth discontinuities and the number of valid returns of the LIDAR, the time required is reduced between 2.5 and 5 times.

2) *Disparity Image Improvement*: Reducing the DSI not only allows a faster computation, but also a reduction of the errors of the final disparity image. To show the improvement



Figure 8. Comparison of disparity images for two frames of the “House” data set. Each subfigure shows the left image of the stereo system (top-left); SRI, where the range color encoding has been changed to show the original gray value obtained from the left image (top-right), the disparity image obtained from the full DSI (bottom-left); and the improved disparity image obtained by using the reduced DSI (bottom-right). The marked regions on the images show the areas that were improved using our proposed algorithm.

achieved, we have computed a final disparity image from a DSI by applying WTA optimization to select the best disparity. The zero-mean sum of square differences with a window size of 15×15 pixels was used as the dissimilarity measure. Since no further refinements are performed, the resulting disparity images are noisy, especially at occluded areas (see Fig. 8). However, our objective here is to show the advantages of using a reduced DSI in comparison to the full DSI. In Section V-D, we refine this raw output to obtain high precision disparity images.

Figure 8 compares the results of standard and reduced DSI using one of our data sets. Most of the errors in the standard method are caused by the lack of texture, repetitive structures, or perspective distortions. These errors are mitigated by using the reduced DSI, providing smoother and more precise disparity images.

B. Path Promotion in DP

In this experiment, we have computed DSIs using the zero-mean sum of squared differences with a window size of 5×5 pixels. No refinements were performed on the obtained disparity images.

Magnitude and Gradient Promotion. Fig. 7 shows the individual contributions of the magnitude and gradient promotion in the DP algorithm. The figure compares the results of WTA, standard DP (Eqs. 1-3), only magnitude promotion DP (Eqs. 4 and 5 with S as defined in Eq. 3), only gradient promotion (Eqs. 1 with S as defined in Eq. 6), and full path promotion (Eqs. 4-6). Observe that the main benefit of the gradient promotion occurs in horizontally slanted regions (e.g., handrail at the left).

Full Path Promotion and Reduced DSI. Fig. 9 compares the results of WTA and DP at two frames of one of the acquired data sets. As it can be seen from the disparity images, both WTA and DP greatly benefit from the LIDAR information, producing smoother disparity images with less false correspondences. The proposed DP with path promotion performs best, providing disparity images with fewer artifacts.

C. Limitations of the Proposed Approach

In general, better stereo results are obtained when LIDAR data is used to guide the search in the disparity space image. Nevertheless, there are some cases in which LIDAR fails to deliver accurate or complete information, leading the stereo algorithm to fail in the estimation of the best disparity. In the example shown in Fig. 10, the LIDAR does not see vehicle’s back windshield. Since the LIDAR is to the right of the camera, it can also measure some of the points on the street that are occluded from the camera point of view. Normally, when predicting the disparity images, these points would be occluded by foreground points from the windshield. Since the foreground data is missing, a wrong prediction of disparities occurs. We are currently evaluating a possible solution to this problem, which is to detect “holes” in the DSI due to missing data and then relax the DP prediction penalty in the vicinity of holes.

D. Final Refinement of Disparity Images

The obtained disparity images can be further refined to reduce artifacts and improve the accuracy of the, up to now, integer stereo disparity computed (Fig. 11). The refinement consists of three steps. First, a median filter is applied to eliminate isolated outliers. Second, a sub-pixel estimation

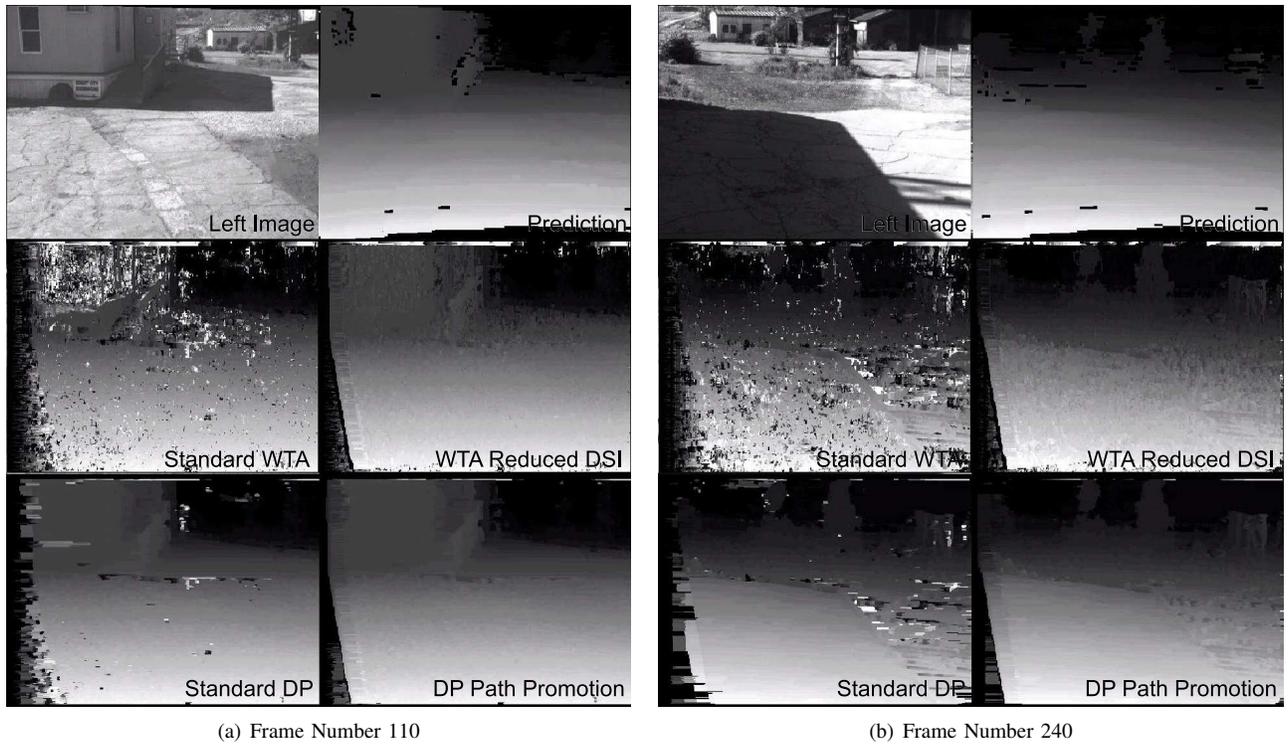


Figure 9. Comparison results of disparity images at two frames of the “House” data set. Top-left: left stereo image; top-right: predicted disparity image from SRI; middle-left: plain WTA optimization results; middle-right: WTA with reduced DSI; bottom-left: standard DP (using Eqs. 1-3); bottom-right: DP with reduced DSI and full path promotion (using Eqs. 4-6).

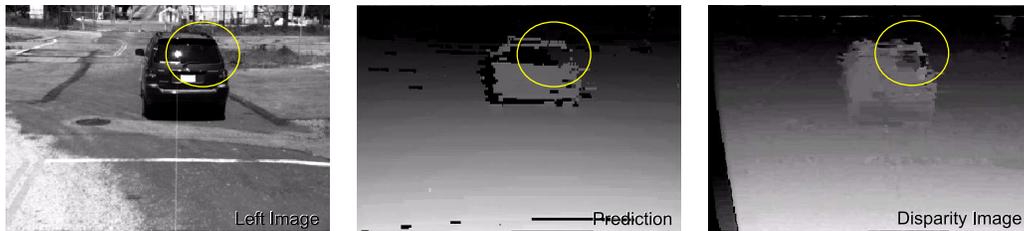


Figure 10. The proposed method can break down when LIDAR data is not measured, such as on the vehicle’s back windshield.



Figure 11. Results of disparity images before the refinement step (second row) and after (third row) for five video frames.

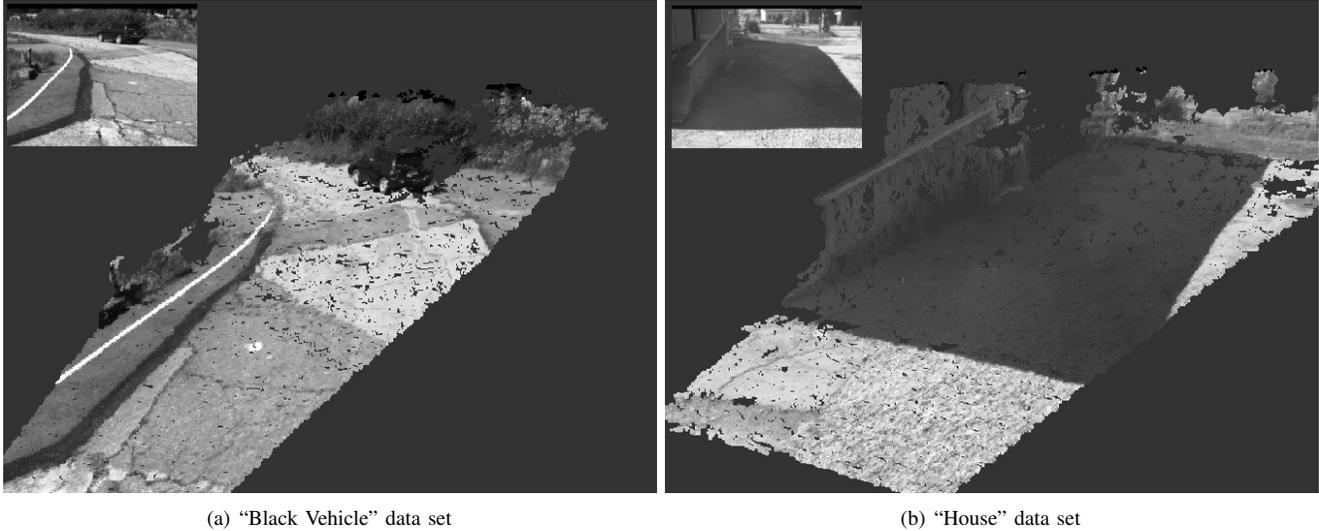


Figure 12. 3D reconstructions obtained from the refined disparity images.

procedure is applied by fitting a parabola to the disparity score and its two neighbors. The sub-pixel disparity is found as the location of the curve where the slope is zero. Third, a left-right-left consistency check is performed by obtaining not only a left, but also a right disparity image, and eliminating the inconsistencies between both views. Figs. 11 and 12 shows three 3D reconstructions obtained with our proposed methods, including this refinement step.

VI. CONCLUSION AND FUTURE WORK

The fusion of LIDAR and stereo leads to a clear improvement of the final disparity image. The computational time is reduced by the appropriate computation of maximum and minimum disparity images. Regrettably, the lack of outdoor sequences with ground truth data prevents us from objectively quantifying the level of improvement achieved. We are planning to generate ground truth data in a controlled environment to address this limitation.

VII. ACKNOWLEDGMENTS

This work was supported by the Agency for Defense Development, Jochiwongil 462, Yuseong, Daejeon, Korea.

REFERENCES

- [1] J. Hackett and M. Shah, "Multi-sensor fusion: a perspective," *Trends in Optical Engineering*, vol. 1, pp. 99–118, 1993.
- [2] P. Rönholm, E. Honkavaara, P. Litkey, H. Hyypä and J. Hyypä, "Integration of laser scanning and photogrammetry," *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 36(Part 3/W52), pp. 355–362, 2007.
- [3] C. Beder, B. Bartczak, and R. Koch, "A combined approach for estimating patchlets from pmd depth images and stereo intensity images," in *Pattern Recognition, 29th DAGM Symposium*, 2007.
- [4] P. Gurram, H. Rhody, J. Kerekes, S. Lach, and E. Saber, "3d scene reconstruction through a fusion of passive video and lidar imagery," in *Applied Imagery Pattern Recognition Workshop, 2007. AIPR 2007. 36th IEEE*, 2007, pp. 133–138.
- [5] K.-D. Kuhnert and M. Stommel, "Fusion of stereo-camera and pmd-camera data for real-time suited precise 3d environment reconstruction," in *IROS, 2006*, pp. 4780–4785.
- [6] K. Nickels, A. Castano, and C. M. Cianci, "Fusion of lidar and stereo range for mobile robots," in *Int. Conf. on Advanced Robotics*, 2003.
- [7] U. Hahne and M. Alexa, "Combining time-of-flight depth and stereo images without accurate extrinsic calibration," *Int. J. Intell. System. Technol. Appl.*, vol. 5, no. 3/4, pp. 325–333, 2008.
- [8] W. F. Oberle and L. Davis, "Toward high resolution, lidar-quality 3-d world models using lidar- stereo data integration and fusion," *Army Research Lab Aberdeen Proving Ground*, Tech. Rep. ARL-TR-3407, 2005.
- [9] C. Mutto, P. Zanuttigh, and G. Cortelazzo, "A probabilistic approach to tof and stereo data fusion," in *3DPVT10*, 2010.
- [10] J. Zhu, L. Wang, R. Yang, and J. Davis, "Fusion of time-of-flight depth and stereo for high accuracy depth maps," in *CVPR*, 2008.
- [11] Y. Kim, C. Theobalt, J. Diebel, J. Kosecka, B. Matusik, and S. Thrun, "Multi-view image and tof sensor fusion for dense 3d reconstruction," in *3DIM*, 2009, pp. 1542–1549.
- [12] S. A. Gudmundsson, H. Aanaes, and R. Larsen, "Fusion of stereo vision and time-of-flight imaging for improved 3d estimation," *Int. J. Intell. Syst. Technol. Appl.*, vol. 5, pp. 425–433, November 2008.
- [13] U. Hahne and M. Alexa, "Depth imaging by combining time-of-flight and on-demand stereo," in *Dynamic 3D Imaging, DAGM Workshop, Dyn3D 2009*, 2009, pp. 70–83.
- [14] Q. Yang, K.-H. Tan, B. Culbertson, and J. Apostolopoulos, "Fusion of active and passive sensors for fast 3d capture," in *IEEE International Workshop on Multimedia Signal Processing*, 2010.
- [15] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, no. 1-3, pp. 7–42, 2002.