

Obstacle Detection during Day and Night Conditions using Stereo Vision

Gijs Dubbelman^{1,2}, Wannes van der Mark¹, Johan C. van den Heuvel¹, Frans C.A. Groen²

Abstract— We have developed a stereo vision based obstacle detection (OD) system that can be used to detect obstacles in off-road terrain during both day and night conditions. In order to acquire enough depth estimates for reliable OD during low visibility conditions, we propose a stereo disparity (depth) estimation approach that uses fine-to-coarse selection in a stereo image pyramid. This fine-to-coarse selection is based on a novel disparity validity metric that reflects the estimation reliability. Dense three-dimensional terrain data is reconstructed from the estimated stereo disparities. In our OD methods, several geometric properties, such as the terrain slope, are inspected to distinguish between obstacles and drivable terrain. This is achieved in a robust and efficient manner by considering the inherent uncertainty in stereo depth and using a hysteresis threshold. A large and varied collection of day- and nighttime images has been used to evaluate the performance of our system. The results show that our methods can reliably detect different types of obstacles in all tested conditions.

I. INTRODUCTION

Unmanned ground vehicles (UGV) can be used for tasks that are considered to be too dangerous or tedious to be performed by humans. Example applications can be found in disaster areas, land mine detection, personnel rescue, construction work, agricultural harvesting and automated transport. Often, such tasks have to be accomplished in difficult terrain during both day and night-time conditions. An autonomous robot vehicle can only perform these tasks safely if it avoids obstacles such as trees, rocks, ravines, people etc. It therefore needs obstacle detection (OD) algorithms that identify obstacles in the path of the vehicle. A considerable research effort has been put into vision based OD for off-road terrain [1], [6], [9]. Unfortunately, all these approaches have only been developed for or tested in day-time conditions. An exception is the work by Owens and Matthies [10], [15] on stereo vision based OD in night-time conditions. Their system relies on a stereo setup of cooled Forward Looking Infra-Red (FLIR) cameras. Instead of using such expensive hardware to acquire sufficient depth estimates for OD, we developed a multi-resolution stereo algorithm that can work with images recorded during low visibility (night-time) conditions by an ordinary stereo camera. It only requires that the scene is dimly lit during the night, in our case by an inexpensive 36 Watt infrared spotlight mounted on our test vehicle, see Fig. 1. The developed stereo algorithm will be explained in the next section. In section III

Electro-Optical Systems¹, TNO Defence Security and Safety, Oude Waalsdorperweg 63, 2509 JG, The Hague, The Netherlands, {gijs.dubbelman, wannes.vandermark, johan.vandenheuvel}@tno.nl

Intelligent Systems Laboratory Amsterdam² (ISLA), University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands groen@science.uva.nl



Fig. 1. RoboJep, TNO's research vehicle for autonomous navigation. The stereo cameras and the infrared spotlight are mounted above the headlights.

the obstacle detection algorithms are described. Section IV presents our benchmark environment and our test results. Our conclusions are given in section V.

II. MULTI-RESOLUTION STEREO ALGORITHM

The first step in our OD system is to reconstruct the three dimensional (3-D) properties of the terrain. This requires that the disparities between corresponding left and right image points are estimated first. In order to improve estimation robustness during night conditions, our approach uses a fine-to-coarse multi-resolution technique, as outlined in Fig. 2. Because of the fine-to-coarse processing, there are no problems with the error propagation usually associated with coarse-to-fine approaches [14]. Input stereo images are assumed to be rectified [2] so that corresponding left and right point reside on the same image lines. In order to create successive levels of the image pyramid, the image size is reduced by 65% repeatedly with bi-linear interpolation. Then, each level in the stereo image pyramid is preprocessed to compensate for radiometric distortion. We compared two preprocessing techniques; the often used Laplacian of Gaussian (LoG) filter [4], [7] and Rank preprocessing, Zabih and Woodfill [17]. After preprocessing, the algorithm estimates a dense disparity map for each level in the stereo image pyramid independently from each other. Then we combine reliable disparity estimates selected from each pyramid level into one disparity image. The selection is based on our disparity validity measure that reflects the disparity estimation quality. In the following sections the stereo algorithm and the disparity validity measure are discussed more thoroughly.

A. Single resolution Stereo algorithm

We use a real-time stereo algorithm, from a framework developed by van der Mark and Gavrila [7], to compute

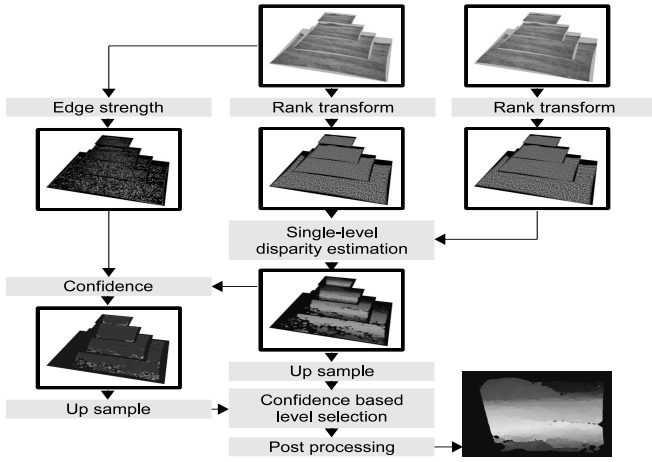


Fig. 2. Stereo algorithm overview.

the disparities for each level of the pyramid. It uses the Sum of Absolute Difference (SAD) of multiple windows to compare left and right points during the disparity search. Only matches with the lowest SAD value are selected, no further search optimization is applied. Sub-pixel accurate disparity estimates are obtained with quadratic interpolation of the SAD values near the found minimum. By performing the disparity search in both the directions left-to-right and right-to-left, the consistency of the estimates can be checked. Finally, in order to remove speckle noise and smooth the remaining estimates, the resulting disparity image is filtered with a blob-filter and a median-filter.

B. Disparity validity measure

The basis of our fine-to-coarse approach is the ability to make a distinction between good and bad disparity estimates. Often, the matching costs computed during the disparity search are used to evaluate the quality of a found match. One such measure is the winner margin, that determines the normalized difference between the best and second (or third) best match. It is used in [4], [13] to remove unclear or ambiguous matches. Another source for determining the matching quality is the signal-to-noise ratio of local regions in the original intensity images [12]. For example, disparity estimation on image regions with low signal to noise ratio is usually prone to errors. Also the disparity estimates themselves give an indication about their correctness because very high disparity differences between adjacent pixels are often caused by faulty estimates. This is the basis for post-processing techniques like blob- and median-filtering [9], [13]. After some initial experiments it was clear that for our purposes matching-cost based validity measures were not performing adequately. This is caused by the fact that these measure ignore the local support for the actual found disparity estimate. The same holds for signal-to-noise ratio methods that can only be used to predict unreliable matches.

We propose a validity measure for disparity estimates that is based on the local differences in the disparity estimates as well as the intensity image signal-to-noise ratio. The first

step is to measure the disparity deviations D in a rectangular shaped window. For this we use Eq. 1 where $d_{u,v}$ denotes the estimated disparity at u, v and $R_{u,v}$ the set of N pixel coordinates from a rectangular window around $d_{u,v}$.

$$D_{u,v} = \frac{1}{N} \sum_{(m,n) \in R_{u,v}} |d_{u,v} - d_{m,n}| \quad (1)$$

For calculating the signal-to-noise ratio we compute the intensity-edge strength for rectangular image regions. We use intensity-edge strength because it gives an impression about the signal-to-noise ratio as well as the presence of object edges in the image. For calculating edge strength we first filter the intensity image with a 1-D normalized derivative of a Gaussian kernel, once in the vertical direction E_y and once in the horizontal direction E_x . The edge strength E is then computed as:

$$E_{u,v} = \sum_{(m,n) \in R_{u,v}} \|(Ex_{m,n} \ Ey_{m,n})^T\| \quad (2)$$

The next step is to find possible faulty estimates. We do this by scaling E and D to the unit interval and subtract them from each other using Eq. 3.

$$V_{u,v} = \frac{E_{u,v}}{E_{max}} - \frac{D_{u,v}}{D_{max}} \quad (3)$$

The idea is that large variations in both local disparity and intensity values only occur near object edges. Erroneous estimates, which also show a large variation in disparity values, are distinct from this because they often occur on object surfaces where the edge response is low. Our measure therefore retains varying estimates near object edges or on textured areas and rejects only those on areas with a bad signal-to-noise ratio.

C. Stereo pyramid level selection

In contrast to the often used coarse-to-fine approaches we use a fine-to-coarse selection scheme. Let d_n be the disparity map resulting from level n in the disparity image pyramid and V_n is its accompanying disparity validity map. First all disparity maps and validity maps are resized to the original resolution. Then the following selection scheme is applied, where λ is an appropriate threshold. For any given disparity estimate at (u, v) in d_0 (the bottom of the image pyramid) it is checked whether its validity is higher than λ . When this is true, the disparity estimate is selected and placed into the output image. Otherwise, the check is performed once again at one level higher in the pyramid. If needed, the check is repeated until the highest level is reached. When the validity value at the highest level is still lower than λ , the pixel is marked as rejected.

III. OBSTACLE DETECTION ALGORITHM

The developed OD system uses geometrical properties derived from the estimated stereo disparities to distinguish between drivable and inaccessible areas in the terrain. Two types of obstacle hazards can be detected by our system. There are the *positive obstacles*, such as tree trunks or steep

sand dunes, that extend out of the ground surface. The other type are the *negative obstacles*, such as ditches or holes, that extend into the ground plane. In order to convert stereo image pixels into 3-D points, we need to compute the real-world distance Z from the estimated disparity value d (given in pixels). If the baseline distance b between the cameras and their focal length f is known, the distance Z can be obtained by triangulation. With this value, the height Y within the camera coordinate frame can be computed from the vertical pixel position u :

$$Z = \frac{fb}{d\Delta v} \quad \text{and} \quad Y = Z \frac{(u\Delta u) - O_u}{f} \quad (4)$$

Here Δu and Δv stand for the vertical and horizontal pixel size respectively. O_u is the vertical image coordinate of the projection of the focal point onto the imaging plane, i.e. optical centre. We transform the coordinates from the camera coordinate frame to the vehicle coordinate frame based on a ground plane assumption. In our approach, the stereo camera parameters f and b , the rectification parameters and the ground plane parameters are obtained from a calibration procedure [3]. We note that we use image column based obstacle detection methods for both positive and negative obstacles. Because the horizontal image coordinate v is the same in each column it is omitted in the following formulas for readability reasons.

A. Positive obstacles

Positive obstacles are detected by applying a hysteresis threshold on the measured terrain slope around image points. The terrain slope is defined as the local difference in height dY divided by the local difference in depth dZ . To measure the difference in height and depth around an image point I_u we can select a image point below it I_{u+S} or above it I_{u-S} within the same image column. Here S is called the step height and controls the scale at which we measure the terrain slope. It is important to measure the slope at the same real-world scale throughout the image. For this we define a granularity G , which is defined in meters. The step-height S in pixels can be computed from the terrain granularity G with formula 5, given the focal length f , the vertical pixel size Δ_u and the estimated depth Z_u of the current image point I_u .

$$S_u = \frac{fG}{f + Z_u \Delta u} \quad (5)$$

The question is whether to use a positive S or a negative S . The answer depends on the local terrain geometry around I_u . In our approach we calculate the slope upward and downward. Then we select that slope that has the maximum absolute value:

$$P' = \frac{Y_{u-S_u} - Y_u}{Z_{u-S_u} - Z_u} \quad P'' = \frac{Y_u - Y_{u+S_u}}{Z_u - Z_{u+S_u}} \quad (6)$$

$$P_u = \arg \max_{p \in \{P', P''\}} (|p|)$$

Once the estimated slope P is calculated for every image point (with known depth) we can apply a hysteresis threshold.

Other obstacle detection systems, [1], [9], [10], often rely on single constant threshold values. The problem with these column based OD approaches is that the terrain slope is measured at the intersection of the *slicing-plane* (plane defined by the optical centre of the camera and the current image column used for slope estimation) and the surface. When the slicing-plane is parallel to the surface normal at the point of intersection the slope will be measured correctly. However, when the slicing-plane deviates from the surface normal the measured slope can be significantly less than the true slope. This is referred to as the “slicing-plane effect”. In [6] the authors propose 3-D clustering of image points as a solution to this problem. To increase the performance of this approach the uncertainty in estimated depth values can be utilized [8]. Unfortunately, achieving real-time frame-rates for these 3-D clustering methods is extremely challenging.

Our assumption is that using hysteresis thresholds is a more computationally efficient alternative for solving the slicing-plane problem. Due to the slicing-plane effect and other disparity estimation artifacts patches of candidate obstacle pixels can be too small to be considered as a true obstacle. By using a hysteresis threshold these patches of candidate obstacle pixels, or seed pixels, can increase in size within the area defined by near-candidate obstacle pixels (grow pixels). Therefore, scattered patches of seed pixels on an obstacle will be grouped together. Hence, the new patches are large enough to be considered as obstacles. When using a hysteresis threshold for positive OD, seed pixels are defined as image points with a non-traversable slope, see Fig. 3. Grow pixels are defined as points with a slope that is almost not traversable by the vehicle. The effect on the false detection rate is kept minimal because grow-pixels, which are not connected to seed pixels, will not be labeled.

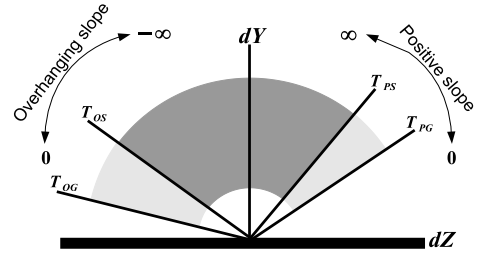


Fig. 3. Hysteresis slope threshold, seed pixels dark gray, grow pixels light gray.

For positive obstacle detection, we use a hysteresis threshold in the following manner. Every pixel with a slope in the range of $[T_{pg} \dots T_{ps}] \cup [T_{og} \dots T_{os}]$ will be marked as a grow pixel, see Fig. 3. All pixels within the range of $[T_{ps} \dots \infty) \cup [T_{os} \dots -\infty)$ will be marked as seed pixels. For all seed pixels we perform binary morphological opening with a small square kernel (3x3) to find consistent patches of seed pixels. The seed pixels left after the morphological opening will be marked as positive obstacles. The pixels that are discarded by this operation are added to the set of grow pixels. Grow pixels will only be marked as a positive

obstacle if there is a 4-connected path of grow pixels to one of the seed pixels. These connected grow pixels can be found efficiently by using region filling.

B. Negative obstacles

Negative obstacles are detected by looking for depth jumps in the range profile of an image column. Based on the estimated depth of an image point Z_u , the chosen terrain granularity G and the height of the camera above the ground C , we can calculate the expected depth jump with:

$$EdZ_u = Z_u \frac{G}{C - G} \quad (7)$$

This expected depth difference can be compared against the estimated depth difference dZ . When calculating the estimated depth difference, the uncertainty in depth estimates should also be considered. The uncertainty $U(Z)$ of distance value Z can be obtained by error propagation of the disparity error εd in the triangulation function of Eq. 4:

$$U(Z) = \frac{\sqrt{2}\varepsilon d Z^2}{fb} \quad (8)$$

It is assumed that the disparity error size is $\pm\Delta v/2$ (despite the sub-pixel estimation). By utilizing this uncertainty we can compute a minimum bound for dZ with:

$$dZ_u = (Z_{u+s_u} - \frac{U(Z_{u+s_u})}{2}) - (Z_u + \frac{U(Z_u)}{2}) \quad (9)$$

For detecting negative obstacles, the ratio between the uncertainty corrected depth jump and the expected depth jump is calculated:

$$N_u = \frac{dZ_u}{EdZ_u} \quad (10)$$

On this ratio a hysteresis threshold is applied. Every pixel with a negative dY and a depth jump ratio in the range of $[T_{ng} \dots T_{ns})$ will be marked as a grow pixel. All pixels with a negative dY and a depth jump ratio in the range of $[T_{ns} \dots \infty)$ will be marked as seed pixels. Again, the seed pixels are processed with morphological opening and region filling is used to find grow pixels that are connected to seed pixels.

C. Grouping and obstacle refinement

The OD system is made less sensitive to disparity estimation artifacts by using depth dependent thresholds on the candidate obstacles. For positive obstacles the threshold is applied to the obstacle's height, for negative obstacles it is applied to its width. In order to measure these dimensions, the initially found object pixels have to be grouped together.

For positive obstacles, grouping of pixels is done by taking both their image coordinates and depth values into account. An obstacle pixel can only be grouped to its 8-connected neighbours. Furthermore, it can only be grouped if the depth difference between the two pixels is smaller than two meters. To prevent obstacle cluttering over the ground plane we perform morphological opening, with a horizontal bar kernel (7x3) on the initial obstacle map. For negative obstacles the following scheme is used. For every vertical image column in

a negative obstacle we only keep that pixel that is closest to the negative obstacle's edge. We find this pixel by searching for the maximum depth jump between a pixel and its upper neighbour. The search range starts at the lowest obstacle pixel and ends at the highest obstacle pixel in the image column extended with the pixel step height S (used during slope estimation). The next step is to segment the negative obstacle map in separate obstacles. In contrast with positive obstacles, grouping of negative obstacle pixels is only based on image coordinates and not on depth values.

We observed that false obstacle detections can be caused by several different types of disparity estimation artifacts. Firstly, by using a fronto-parallel aggregation window fronto-parallel surfaces are favoured [16]. Applying such a fronto-parallel aggregation window for ground plane pixels causes phantom vertical plane artifacts. The severeness of these artifacts increases with range. This is due to the fact that the relative resolution ($pixels/m^2$) inside an aggregation window decreases with depth. The second artifact is caused by sub-pixel interpolation and is referred to as the range-ripple effect [5]. It is caused by the fact that sub-pixel interpolation favors pixel accurate estimates. It is hard to find a closed form model for the effects that these and other disparity estimation artifacts have on obstacle detection. Therefore, the following exponential function is used for our depth dependent thresholds.

$$T(O_z) = e^{\frac{O_z}{\alpha}} - \beta \quad (11)$$

Here, β sets the minimal obstacle size ($1 - \beta$) at zero range, α controls the range at which the threshold curve reaches $(e - \beta)$ meters and O_z is the average range to the obstacle. In order to get obstacle size measurements that are less sensitive to outliers in the 3-D estimates we measure the projected size in pixels. To convert the threshold curve given in meters to pixels the following formula is used.

$$T'(O_z) = \frac{fT(O_z)}{\Delta y O_z} \quad (12)$$

Finally, an obstacle passes the threshold when its pixel size O'_{size} (height for positive obstacles and width for negative obstacles) is larger than the threshold based on its range:

$$T'(O_z) \leq O'_{size} \quad (13)$$

After the candidate obstacles are filtered based on the depth dependent thresholds we apply some post processing techniques. These post processing methods mainly increase the visual quality of the obstacle map. Negative obstacle pixels are dilated with a small (9x9) diamond shaped kernel. This causes the one-pixel thick lines to become more visible in the image. For positive obstacles we look at their vertical image columns. Every pixel that has a pixel above and below it in the same image column from the same object, will also be marked as belonging to that positive obstacle.

IV. RESULTS

In order to evaluate the proposed algorithms, we recorded stereo image sequences in varied off-road terrain during day

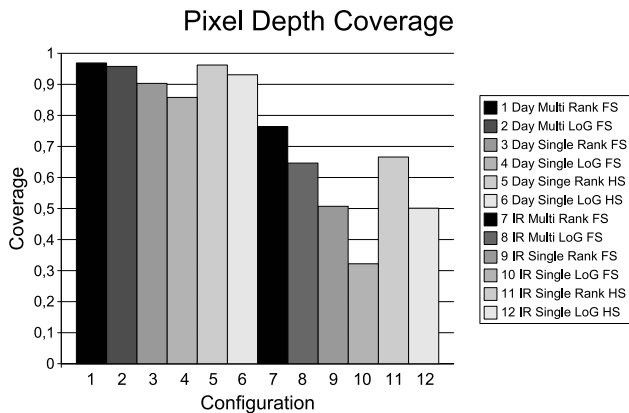


Fig. 4. Depth coverage

and night conditions. During the night the scene was illuminated with a 36 Watt IR emitter mounted on the vehicle. From these recordings we manually selected 140 daytime and 140 night-time images. Time synchronization was used to associate GPS data to each recorded camera image. By matching the GPS data and by visual inspection, we selected for every frame in the daytime dataset a comparable frame in the night-time dataset. All 140 images from both our daytime and night-time dataset have been manually labeled to obtain groundtruth data, see Fig. 9. The labeling is pixel based and consists out of four classes: the positive obstacle class, the negative obstacle class, the drivable class and the ignore class. We evaluated the obstacle detection itself as well as the stereo depth reconstruction performance using a wide range of parameter settings. We experimented with both our novel multi-resolution (Multi) technique (section II) and a single-resolution (Single) method (section II-A). For the single-resolution approach we tested the performance using the full (FS 640x480) and half (HS 320x240) resolution. Furthermore we experimented with LoG and Rank preprocessing and measured their effect for the day-time and night-time datasets. For obstacle detection we experimented with 40 different parameter settings. The step height S was varied in the range of 0.15 m up to 0.65 m. We also tried several hysteresis configurations.

A. Depth coverage

The depth coverage of our tested methods is presented first, see Fig. 4. The depth coverage is the percentage of pixels which are not rejected. Rejection is based on the left-to-right consistency check and for the multi-resolution approaches also on their validity. Evaluation was performed using the 140 day-time (Day) and 140 night-time (IR) images. It can be clearly seen that, as a preprocessing step, the Rank transform outperforms LoG filtering. Especially during night time conditions the Rank transform shows an increase of at least 10% compared to LoG filtering. Another interesting aspect is the performance gain during night conditions due to the novel multi-resolution method. Again a 10% increase in depth coverage is achieved between the multi-resolution and

single-resolution approaches. By using the single-resolution approach at full image size we do not reach an acceptable coverage level during night-time conditions. Most likely this is due to the fact that the matching windows do not contain enough distinctive intensity variation. An acceptable coverage level on half resolution night-time images can only be achieved with the single-resolution approach if Rank preprocessing is used. Some example disparity maps for different estimation methods and configurations are shown in Fig. 5.

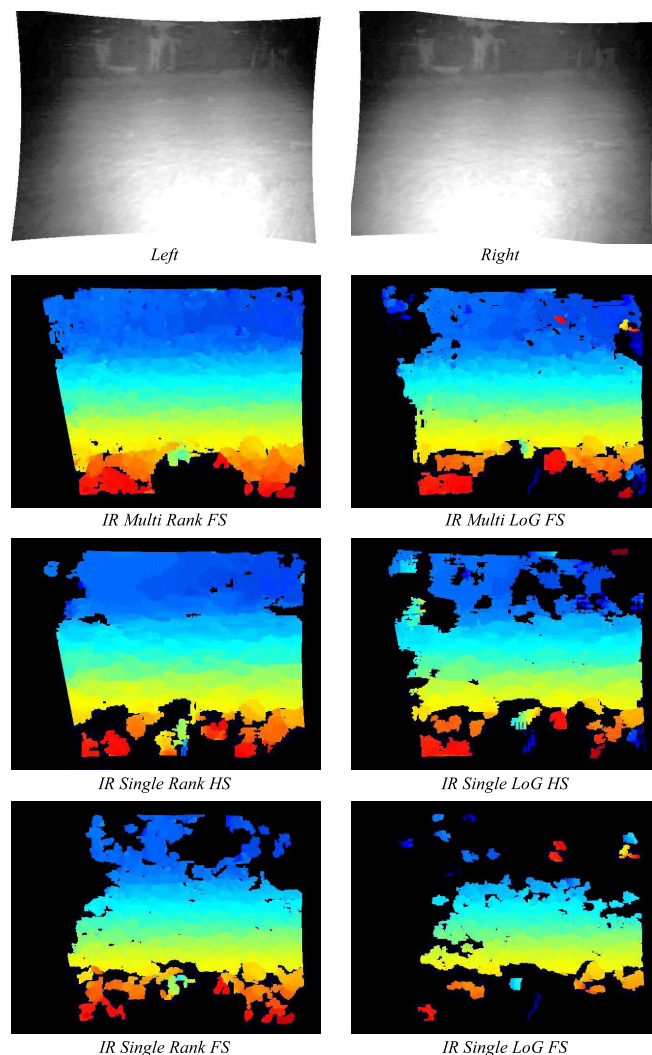


Fig. 5. Example disparity maps for different estimation methods and configurations. Disparity is color coded from large (red) to small (blue). Black indicates that no reliable estimate was found.

B. Stereo pyramid level selection

Another aspect of our multi-resolution approach is the possible decrease in depth uncertainty. This is due to the fact that we can use higher resolutions while at the same time keep adequate depth coverage. Whereas the single-resolution approach only reaches adequate depth coverage when using half of the original image size. Using half of the initial resolution effectively doubles the pixel size and thus

increases depth uncertainty. In Fig. 6 we present the influence of the multi-resolution approach on depth uncertainty. The advantage of our approach is that it will take estimates from different resolutions. Some estimates will originate from the base of the image pyramid (highest resolution) having minimum uncertainty. Other estimates will originate from lower resolutions that contain more uncertainty. The question arises how much each pyramid level contributes to the disparity map. For several configurations we plotted the percentage of pixels coming from the four different levels in the disparity image pyramid. Again the evaluation was based on the 140 day-time and 140 night-time images in our dataset. During

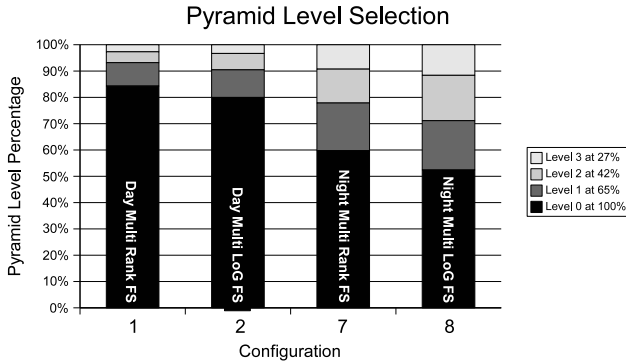


Fig. 6. Disparity pyramid level selection.

day-time conditions the majority of the estimates come from the base of the stereo image pyramid (maximal resolution) and consequently have minimal depth uncertainty. During night-time conditions the base of the stereo image pyramid contributes 60% for Rank preprocessing and 52% for LoG preprocessing. This is consistent with the coverage levels of the multi-resolution and single-resolution configurations at full image size, see Fig. 4.

C. Positive obstacle detection

We propose a positive obstacle evaluation method not based on the pixels themselves but on the surface that those pixels represent. Due to the projection of objects on the image plane, an object of particular size will consist of fewer pixels at large range than at close range. Hence, by using a pixel based evaluation scheme we would skew the results in favour of objects near to the vehicle. By using a surface based evaluation scheme the performance is unbiased, because every pixel is weighted according to the real-world surface it represents. The estimated depth of a pixel together with the camera parameters and some simplifying assumptions based on its groundtruth label, are used to calculate the surfaces the pixel represents. By doing this for every pixel (with known depth) we can compute the groundtruth surface of all positive obstacles as well as the surface of the drivable terrain in our datasets. The groundtruth surfaces are computed according to the day-time dataset. For a given estimated obstacle map we can distinguish between true positive obstacle pixels and false positive obstacle pixels based on the groundtruth label maps. Because the surface represented by the pixels

can be calculated, we can also measure the percentage of correctly classified obstacle surface as well as the percentage of wrongly classified driveable terrain. For each of the 40 parameter settings we plot a marker in the following graphs. Every marker shows the obtained surface percentage off correctly labeled obstacles and the surface percentage of wrongly labeled driveable terrain measured over the whole dataset.

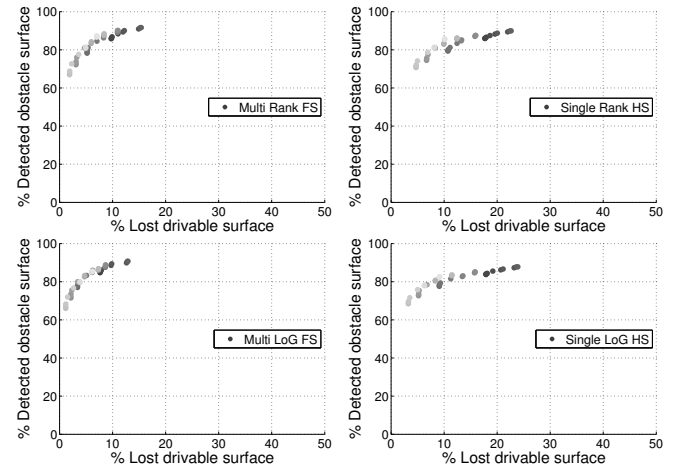


Fig. 7. Day-time positive obstacle detection up to 50 m. Every marker presents the performance over the whole day-time dataset with one of the 40 tested parameter settings.

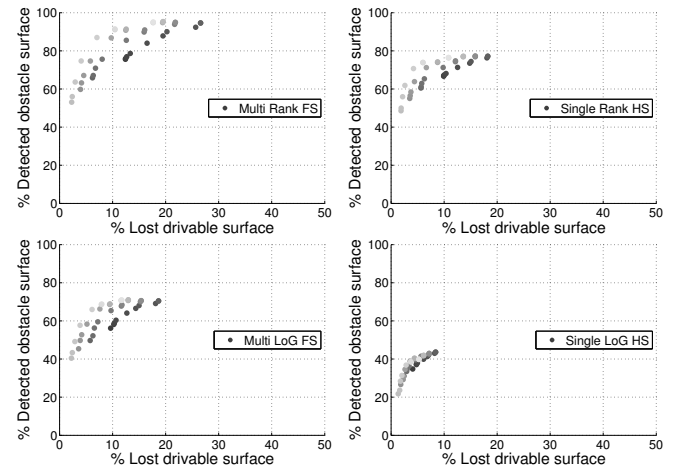


Fig. 8. Night-time positive obstacle detection up to 25 m. Every marker presents the performance over the whole night-time dataset with one of the 40 tested parameter settings.

Fig. 7 shows the positive OD performance on our day-time dataset for different disparity estimation and positive OD configurations. Evaluation was performed for terrain up to 50 m away from the vehicle. It should be noted that these range values are also obtained from the estimated stereo disparities. Nevertheless, the results are encouraging. The influence of the different disparity estimation configurations on daytime OD performance is minimal. The single-resolution method has a slightly worse false detection rate

due to increased disparity dilation effects. This is caused by using the same matching window size, i.e. 11x11, on smaller image resolutions i.e. 320x240. Fig. 8 presents the positive OD performance on our night-time dataset. Because performance is unacceptable if all distances up to 50 m are considered we evaluated up to 25 m. In the plots we clearly see the influence of the tested depth estimation techniques on night-time positive OD. Both our multi-resolution approach and Rank preprocessing have a noticeable positive effect on OD performance. The performance gain between depth estimation configurations in true detection rate is about the same as the gain in depth coverage for the given configurations. This indicates that the extra depth information found using the multi-resolution approaches has sufficient quality to increase the detection of positive obstacles. It can be observed that only the configuration that uses Rank preprocessing and our multi-resolution approach reaches acceptable performance levels for obstacles up to 25 m during night-time conditions. The influence of the OD parameters is not clear from these images. However we observe that using a larger step-height reduces the false detection rate. Unfortunately it also reduces the true detection rate. When larger step-heights are chosen, using hysteresis thresholds can boost the true detection rate while keeping the false detection rate fairly stable. We observed that by using a terrain granularity G of 0.45m and modest hysteresis threshold values i.e. $grow = [0.45 \dots 0.75] \cup [-0.10 \dots -0.10]$ and $seed = [0.75 \dots \infty] \cup [-0.10 \dots -\infty]$, optimal results are achieved. The effect of using hysteresis threshold on the slicing-plane effect is shown in Fig. 12. We clearly see that due to the slicing-plane effect an obstacle can be inconsistently classified. By using hysteresis threshold classification becomes more consistent over the obstacle's surface.

D. Negative obstacle detection

Using our surface based evaluation approach requires a large evaluation dataset. Otherwise depth estimation errors can influence the evaluation significantly. Unfortunately, our dataset only contains one sequence with a single negative object. This is why we considered using another method to measure the performance on negative obstacles. The negative OD performance is based on the percentage of a negative obstacle's width that is correctly classified. Here, we measure the obstacle width in pixels. For the false detection rate of our negative OD method we use an approach similar to that of positive obstacles. However, before the horizontal surface of false negative obstacles is computed, the false detections are dilated with a rectangular kernel. This is done because false negative obstacles can occur as small lines only a few pixels in length scattered throughout the image. By dilating these pixels, the measure reflects the negative effect on the vehicle's expected performance more realistic.

Negative obstacle are more challenging than positive obstacles, as can be seen in Fig. 11. While adequate performance can be achieved for distances up to 10 m during the day, the false detection rate becomes unacceptable at larger

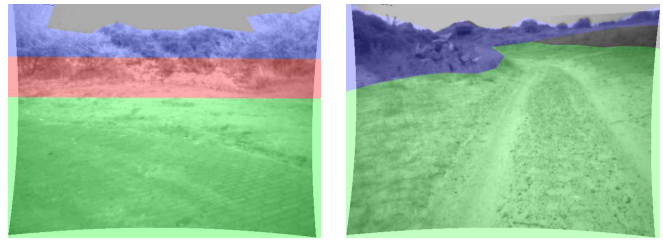


Fig. 9. Example groundtruth labelmaps. The color green indicates the ground surface, blue positive obstacles and red negative obstacles.

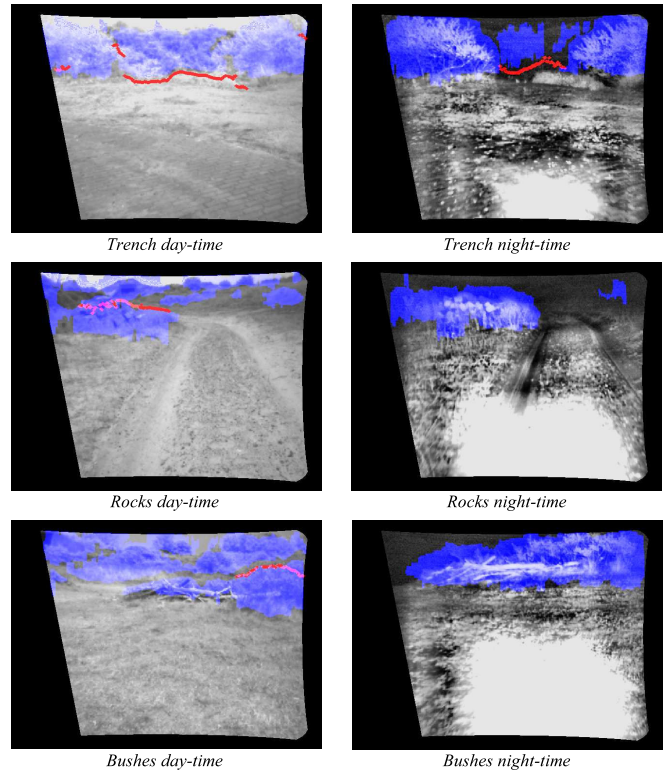


Fig. 10. Day- (left side) and night-time (right side) OD results. The color blue indicates positive obstacles and red negative obstacles. Because, no ground plane assumption is made during OD, negative obstacles can also appear on top of positive obstacles.

distances and during night-time conditions. Negative obstacle detection is considered one of the most challenging topics for autonomous navigation. Many systems rely on narrow field of view camera's mounted on a pan tilt unit. Thermal imaging has been proposed as an alternative by Matthies and Rankin [11]. However, even when using the just mentioned hardware results are limited. Therefore, we are not surprised by our results. Having said that, at close range our system is able to detect negative obstacles with reasonable reliability.

V. SUMMARY AND CONCLUSIONS

We have presented a stereo vision based approach to obstacle detection in off-road terrain during both day and night conditions. Obstacle detection (OD) at night requires robust depth estimation which we achieve by our novel fine-to-coarse multi-resolution disparity estimation technique.

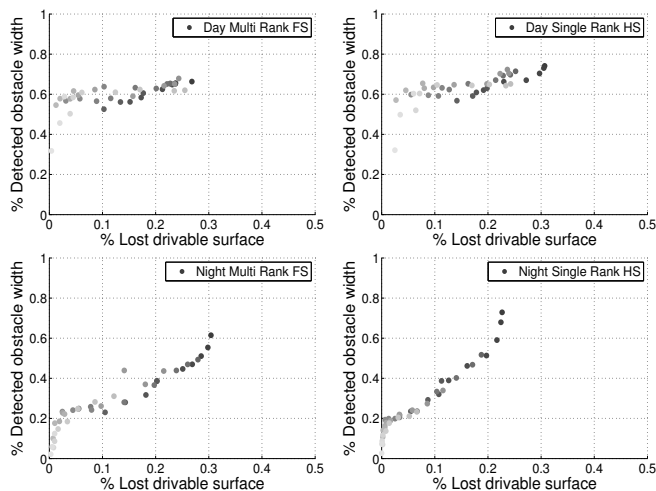


Fig. 11. Day- and Night-time negative obstacle detection up to 10 m. Every marker presents the performance over the whole day/night dataset with one of the 40 tested parameter settings.

Our fine-to-coarse approach does not suffer from traditional coarse-to-fine error propagation and is based on the ability to make a distinction between good and bad stereo matches. Therefore, an important contribution is our disparity validity measure. Which, in contrast to traditional matching-cost based measures, can be used to guide the disparity selection process. LoG filtering is frequently used as a preprocessing step in approaches from the literature, our tests however suggest the Rank transform might be more appropriate. The detection of positive and negative obstacles can be done robustly by depth dependent slope measurement and obstacle thresholds. We introduced using hysteresis thresholds as a powerful and efficient tool for boosting image-column based OD performance. Furthermore, it can be used to reduce the damaging effect of the slicing-plane problem which plagues image-column based OD approaches.

ACKNOWLEDGMENT

This research was sponsored by TNO Defence, Safety and Security. The authors are grateful to Marinus Maris and Han van Bezooijen for their assistance.

REFERENCES

- [1] P. Bellutta, R. Manduchi, L. Matthies, K. Owens, and A. Rankin, "Terrain perception for DEMO III," in *IEEE Intelligent Vehicles Conference*, 2000, pp. 326–331.
- [2] A. Fusiello, E. Trucco, and A. Verri, "Recification with unconstrained stereo geometry," in *British Machine Vision Conference*, 1997.
- [3] J. Heikkilä and O. Siven, "A four-step camera calibration procedure with implicit image correction," in *IEEE Computer Vision and Pattern Recognition Conference*, 1997, pp. 1106–1112.
- [4] H. Hirschmüller, P. Innocent, and J. Garibaldi, "Real-time correlation-based stereo vision with reduced border errors," *IJCV*, vol. 47, no. 1-3, pp. 229–246, 2002.
- [5] W. Kim, A. Ansar, R. Steele, and R. Steinke, "Performance analysis and validation of a stereo vision system," in *IEEE Conference on Systems, Man and Cybernetics*, vol. 2, 2005, pp. 1409–1416.
- [6] R. Manduchi, A. Castano, A. Talukder, and L. Matthies, "Detection and terrain classification for autonomous off-road navigation," *Autonomous Robots*, vol. 18, pp. 81–102, 2005.

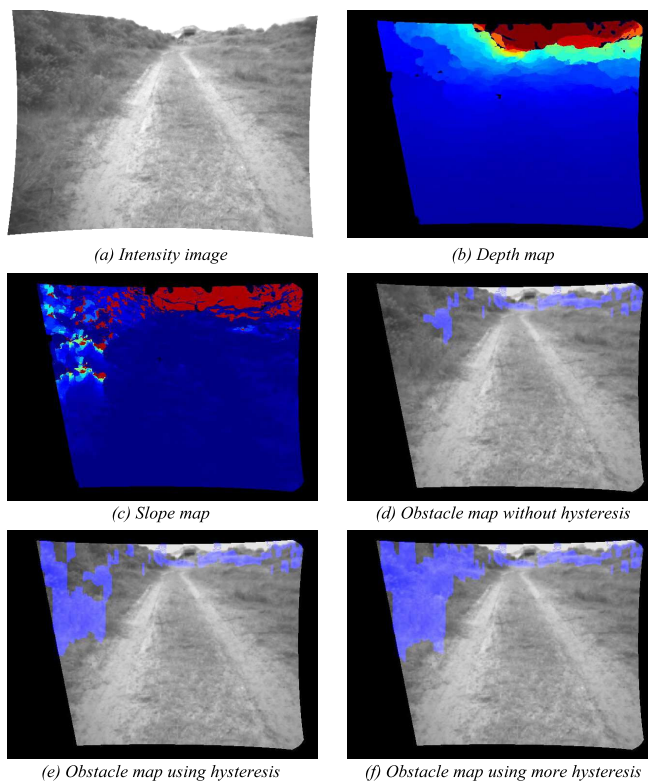


Fig. 12. This image shows the effect of using hysteresis thresholds in relation with the slicing-plane effect. (a) Intensity image. (b) Depth map (blue=near, red=far). (c) Slope-map (blue=flat, red=steep). (d) OD map without using hysteresis. (e) OD map using hysteresis. (f) OD map using more hysteresis.

- [7] W. v. d. Mark and D. Gavrila, "Real-time dense stereo for intelligent vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 7, no. 1, pp. 38–50, 2006.
- [8] W. v. d. Mark, J. v. d. Heuvel, and F. Groen, "Stereo based obstacle detection with uncertainty in rough terrain," in *IEEE Intelligent Vehicle Symposium*, June 13-15 2007, pp. 1005–1012.
- [9] L. Matthies, A. Kelly, T. Litwin, and G. Tharp, "Obstacle detection for unmanned ground vehicles: a progress report," in *IEEE Intelligent Vehicles Conference*, 1995, pp. 66 – 71.
- [10] L. Matthies, T. Litwin, K. Owens, and et al., "Performance evaluation of ugv obstacle detection with ccd/flir stereo vision and lidar," in *IEEE Workshop on Perception for Mobile Agents*, 1998.
- [11] L. Matthies and A. Rankin, "Negative obstacle detection by thermal signature," in *IEEE Intelligent Robots and Systems Conference*, 2003, pp. 906–913.
- [12] H. Moravec, "Towards automatic visual obstacle avoidance," in *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, 1977, p. 584.
- [13] K. Mühlmann, D. Maier, J. Hesser, and R. Männer, "Calculating dense disparity maps from color stereo images, an efficient implementation," *IJCV*, vol. 47, no. 1-3, pp. 79–88, 2002.
- [14] M. O'Neill and M. Denos, "Automated system for coarse-to-fine pyramidal area correlation stereo matching," *Image and Vision Computing*, vol. 14, no. 3, pp. 225–236, 1996.
- [15] K. Owens and L. Matthies, "Passive night vision sensor comparison for unmanned ground vehicle stereo vision navigation," *IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, p. 59, 1999.
- [16] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *IJCV*, vol. 47, pp. 7–42, 2002.
- [17] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *ECCV*, vol. 2, 1994, pp. 151–158.