Nonparametric Technique Based High-Speed Road Surface Detection

Meiqing Wu, Siew-Kei Lam, *Member, IEEE*, and Thambipillai Srikanthan, *Senior Member, IEEE*

Abstract—It has been well recognized that detecting road surface in a realistic environment is a challenging problem that is also computationally intensive. Existing road surface detection methods attempt to fit the road surface into rigid models (e.g., planar, clothoid, or B-Spline), thereby restricting to road surfaces that match specific models. In addition, the curve-fitting strategies employed in such techniques incur high computational complexity, making them unsuitable for in-vehicle deployments. In this paper, we propose an efficient nonparametric road surface detection algorithm that exploits the depth cue. The proposed method relies on four intrinsic road scene attributes observed under stereo geometry and has been shown to reliably detect both planar and nonplanar road surfaces efficiently. Extensive evaluations are performed on three widely used benchmarks (i.e., enpeda, KITTI, and Daimler), encompassing many complex road scenarios. The experimental results show that the proposed algorithm significantly outperforms the well-known techniques both in terms of detection accuracy and runtime performance.

Index Terms—Advanced driver assistance systems, depth, non-parametric, obstacle, road surface, V-disparity.
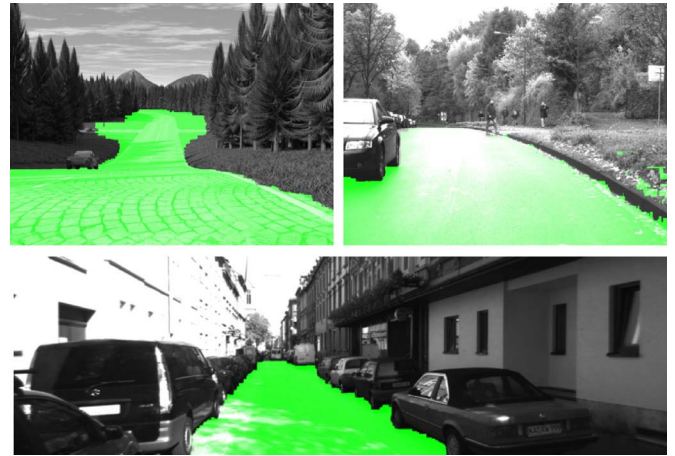

Fig. 1. Three sample results of road surface detection using the proposed approach. (Top row) Two different scenarios with undulating hills. (Bottom row) Urban planar road scenario.

I. INTRODUCTION

THE task of detecting road surface is the first key step toward the realization of autonomous vehicle [1], [2]. Road refers to the drivable area that allows vehicle to proceed without encountering obstructions that would prevent the onward traversal. Road surface detection in realistic road scenarios faces a myriad of challenges. The road surface can be planar or nonplanar (e.g., uphills, downhills, and undulating hills). The shape of the roads can also vary (e.g., straight or curved). In addition, the presence of crowded objects, cluttered background, varying illumination conditions, and dappled shadows make the road surface detection process extremely challenging. In order to deal with such high variability in road scenes, many cues such as color, texture, and depth are often exploited for detecting road surfaces.

In this paper, we propose a nonparametric road surface detection algorithm that relies on the depth cue only. Unlike existing methods, the proposed road surface detection algorithm does not fit the road surface into fixed mathematical models. Instead, it relies on simple but effective strategies that are based on the following four special attributes present in realistic road conditions.

1) The road surface extends away from the camera, and obstacles are often in upright position.
2) The road surface constitutes the major portion of the scene after the large obstacles (e.g., cars, buildings, and trees) are removed, particularly in the vicinity of the vehicle.
3) Due to the perspective projection effect of the camera imaging process, road regions that are further away from the camera will be located in the higher portion of the captured image.
4) The distance of an obstacle point from the camera should be lesser than or equal to that of a road point if they are both projected onto the same row in the captured image.

By formulating these observations, the proposed algorithm is able to work with highly dynamic road scenarios at low computational complexity. Some sample results of road surface detection using the proposed approach are shown in Fig. 1.

This paper is organized as follows. Section II reviews the existing works in road surface detection. Section III introduces the stereo model and the salient road scene attributes that are observed under this model. Based on these observations, the proposed road surface detection algorithm is presented in Section IV. In Section V, we present the experimental results to demonstrate that the proposed algorithm outperforms the baseline methods in various complex road scenarios. Finally, we conclude this paper in Section VI.

II. RELATED WORKS

Road color and texture are the two important cues exploited for monocular vision-based road surface detection. Thorpe *et al.* [2] classify the image points as "road" or "non-road" using multiclass adaptive color classification. The work in [3] estimates the road shape by extracting lane markings. However, lane markings are not present in many realistic road scenarios, and hence, the method in [3] is only applicable to specific road scenes. The approach in [4] detects road surfaces by combining the results of road boundary estimation based on intensity image and road-area segmentation based on color image. References [5] and [6] estimate the vanishing points of roads by exploiting the texture cue first and then localizing the road boundary using the color cue. The work in [7] proposes a road detection algorithm by combining low-level, contextual, and temporal cues in a Bayesian framework. However, this approach imposes a high computational cost.

There is also a large body of work that exploits the depth cue for road surface detection. By restricting the problem to planar roads, the work in [8] finds that the corresponding longitudinal road profile will be projected as a diagonal straight line in the v-disparity image. Curve-fitting techniques, e.g., Hough transform [8], Radon transform [9], and linear regression scheme [10] are then adopted to extract the straight line. Instead of working in v-disparity space, [11] works in the Euclidean space and detects the road surface by fitting the 3-D road data points into a plane using RANSAC-based least squares approach.

Unfortunately, the above techniques restrict the problem to handle only planar roads, hence limiting their applicability in many real-world scenarios. Road surfaces are often highly unstructured due to uphills, downhills, undulating hills, road speed bumps, etc. Recently, researchers have been increasingly shifting their efforts to deal with the nonplanar road geometry. In addition to taking care of the planar road, [8] also models the nonplanar road surface as a succession of parts of oblique planes. The corresponding longitudinal road profile is formulated as a piecewise linear curve. Based on this model, a global road profile is extracted in [12], and a classification and propagation operation is performed to refine the road profile. Reference [13] approximates the road surface as a quadratic model, whereas [14] fits the road surface into a clothoid model. However, all of the aforementioned techniques allow for road slope changes in only one direction [15]. This motivates the work in [15] to represent the road surface as a general parametric B-spline curve. However, it is difficult to determine the surface parameters.

Our work belongs to the category that exploits the depth information for road surface detection. Unlike the existing works that fit the road surface into fixed mathematical models, the proposed method detects road surface by exploiting their intrinsic attributes under a stereo geometry. By mathematically formulating the observed road scene attributes, we are able to devise a simple but efficient nonparametric road surface detection algorithm that is applicable to both planar and nonplanar road scenarios.
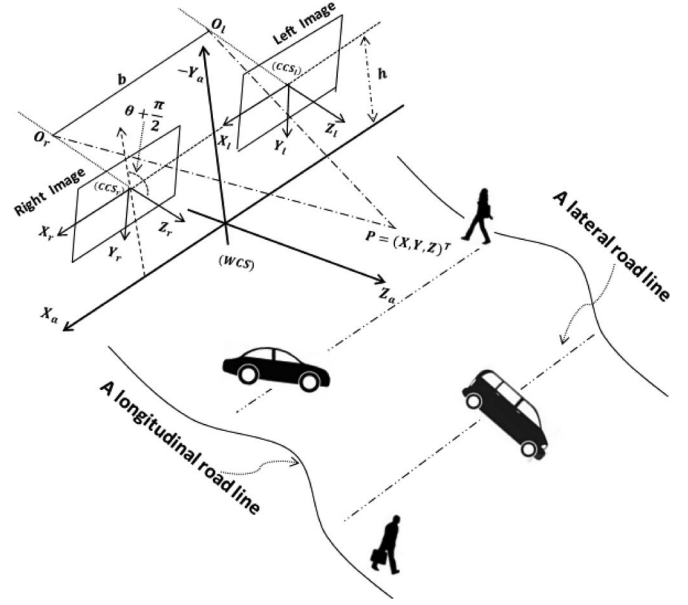

Fig. 2. Stereo model (from [8]) and an example of a road scenario with undulating hill.

III. OBSERVATIONS OF SPECIAL ROAD SCENE ATTRIBUTES UNDER STEREO GEOMETRY

In this section, we will first describe the stereo geometry and the corresponding u-v disparity images briefly. We will then discuss four important road scene attributes observed under the stereo geometry. The observed road scene attributes serve as the mathematical foundations of the proposed algorithm presented in Section IV.

A. Stereo Geometry

A parallel stereo rig is set up as in [8]. As shown in Fig. 2, a scene point $\boldsymbol{P}(X, Y, Z)$ will be projected onto an image point $\boldsymbol{p}(u, v)$. Assume that θ is the cameras' pitch angle; h is the height of the cameras above the ground; b is the stereo baseline; f is the focal length measured in pixels; (u_0, v_0) is the image coordinate of the projection of the optical center of the camera; i indicates left camera or right camera; and $\varepsilon_l = -1$ and $\varepsilon_r = 1$. Then the coordinate (u, v) is given as

$$u_i = u_0 + \frac{fX - \varepsilon_i \frac{b}{2} f}{(Y + h)\sin\theta + Z\cos\theta} \tag{1}$$

$$v = v_0 - f\tan\theta + \frac{f(Y+h)}{\cos\theta\left[(Y+h)\sin\theta + Z\cos\theta\right]}. \tag{2}$$

The disparity value $\Delta = u_l - u_r$ is given in

$$\Delta = \frac{fb}{(Y+h)\sin\theta + Z\cos\theta}. \tag{3}$$

A map containing the disparity value for each pixel is called the disparity map.

A relationship between Δ, v, and Y is given as

$$\Delta = \frac{b(v\cos\theta - v_0\cos\theta + f\sin\theta)}{(Y+h)}. \tag{4}$$
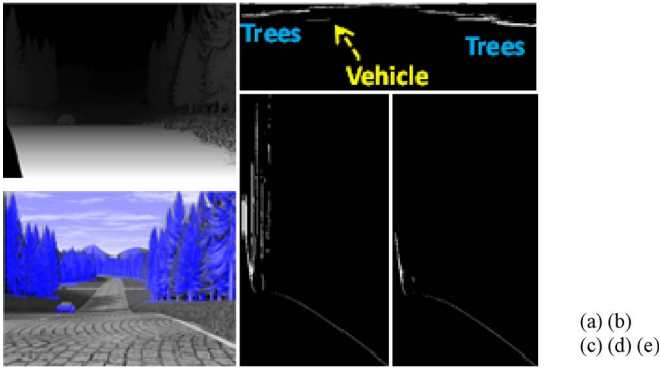
Fig. 3.   (a) Original disparity map and (b) and (d) corresponding u-disparity image and v-disparity image. The obstacles in (c), shown in blue, are identified in the crude obstacle removal step, and the new v-disparity image (after removing obstacles) is shown in (e).

Note that when θ is small enough, $\sin \theta \approx 0$ and the term $(Y + h) \sin \theta$ can be ignored. We assume a small θ in this paper. In addition, we assume that the cameras are installed such that the roll angle is negligible as in [14], [16], and [17].

B. U-V Disparity Images

The concepts of u-disparity image and v-disparity image are first proposed in [8]. V-disparity image is obtained by accumulating the points with the same disparity in the scan line of the disparity map. The v-disparity image provides a side-view projection of the 3-D scene. U-disparity image, on the other hand, accumulates the points with the same disparity in a columnwise manner and therefore provides a top-view projection of the scene. Fig. 3(b) and (d) illustrates the u-disparity image and v-disparity image for the disparity map in Fig. 3(a).

C. Special Road Scene Attributes

The example of a road scene depicted in Fig. 2 shows a challenging road scenario, i.e., nonplanar road with undulating hill and dynamic obstacles on the road. When the road scene is mapped to the stereo coordinate system, as described in Section III-A, several intrinsic road scene attributes in the u-v disparity images are identified, which will enable us to distinguish the road surface from the obstacles. The following describes these observed attributes.

1) A longitudinal road line is defined as a set of road points that have the same X value but varying Z values. According to (1) and (3), a longitudinal road line will fall into distributed regions in the u-disparity image. On the other hand, up-right obstacle points with the same X and Z values will converge onto the same position in the u-disparity image, therefore producing peak regions. Fig. 3(b) clearly illustrates this concept.

2) A lateral road line is defined as a set of road points that have the same Z value but varying X values. In most cases, the points along the same lateral road line have the same value of Y or values of Y that are very close to each other. According to (2) and (3), the lateral road line will be projected onto the same v and associated with the same Δ, i.e., they will converge onto the same position in

the v-disparity image. When a vehicle is moving forward, the road surface will occupy a major part of the scene, particularly in the vicinity of the vehicle. This implies that the point with maximum intensity value for each row in the v-disparity image is very likely to be the projection of the points of the corresponding lateral road line. However, this property will be violated when there are many obstacles (particularly large obstacles) on the road. In order to increase the confidence that the peak regions in the v-disparity image will correspond to the lateral road lines, the obstacles can be first removed from the disparity map prior to generating the v-disparity image.

3) According to (2) and (3), larger values of Z will lead to smaller values of v and Δ. This explains the following phenomenon: Due to the perspective projection effect of the camera imaging process, road points that are farther away from the camera will be projected onto the higher part of the captured image. In addition, they will have smaller disparity values.

4) According to (4), when two points are projected onto the same row v, the point with larger value of Y will have smaller Δ. This implies that when a road point and an obstacle point are projected onto the same row v, the Z value of the obstacle point is smaller than that of the road point. In addition, the disparity value corresponding to the obstacle point is larger than the disparity value for the road point.

IV. PROPOSED ALGORITHM

Existing methods have mainly concentrated on fitting the shape of road surface into a specific model. We argue that it is very difficult to find a single model that can accommodate all road scenarios since the environment is highly dynamic. As such, we have developed a novel road surface detection algorithm that is based on the road scene's intrinsic attributes under a stereo geometry, as discussed in Section III. In this section, we will show how these attributes are combined and formulated into an efficient road surface detection algorithm. The proposed method consists of four steps: 1) crude obstacles removal; 2) longitudinal road profile extraction; 3) determination of the horizon line; and 4) road surface extraction. The only input required by the proposed algorithm is a dense or semidense disparity map.

A. Crude Obstacle Removal

This step serves as a preprocessing step to facilitate road profile extraction that will be discussed in Section IV-B. We resort to the u-disparity image for fast obstacle removal. As explained in Section III-C, obstacles, particularly large obstacles, correspond to peak regions in the u-disparity image. In order to identify these peaks in the u-disparity image, we simply apply a thresholding operation with a threshold denoted as *threshold_largeobj*. If the intensity value of a point in the u-disparity image is higher than *threshold_largeobj*, the point is labeled as obstacle point. A new disparity map is then generated after removing the obstacles from the original disparity map. It
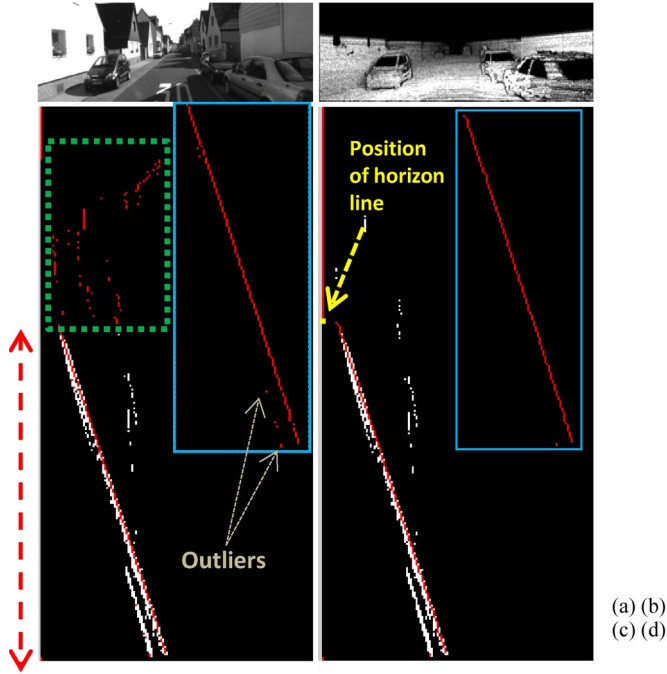
Fig. 4. (a) Captured image and (b) its corresponding disparity map with enhanced contrast. (c) Initial road profile (see the blue inset for enhanced visualization of the range indicated by the red vertical dotted line). (d) Refined road profile (see the blue inset for enhanced visualization).

is noteworthy that this preprocessing stage is not intended at finding all the obstacle pixels present in the disparity map but rather as a means to reduce the difficulty of extracting the road profile. It can be observed in Fig. 3(c) that the crude obstacle removal step removes the sky and trees but retains the short grass. This is, however, sufficient for increasing the accuracy of road profile extraction, which will be discussed in the following section.

B. Longitudinal Road Profile Extraction

In this step, the longitudinal road profile is extracted based on the v-disparity image, which is generated from the new disparity map after removing the large obstacles. As described in Section III-C, the points with maximum intensity value for each row in the v-disparity image are very likely to correspond to the projections of the lateral road lines after removing the obstacles. These points constitute the initial road profile. Fig. 4(c) shows an example of the initial road profile for a given disparity map.

In our experiments, we find that the majority of the initial road profile points do correspond to the actual road profile, particularly in the vicinity of the vehicle. Nevertheless, there can be exceptions due to two reasons. The first reason is that the higher portion of the initial road profile may not correspond to the physical lateral road lines. For example, in Fig. 4(c), the green rectangle corresponds to the higher region of the captured image that does not contain the road. Another reason is due to the highly unstructured road scenes or the noisy and erroneous input disparity map. The initial road profile therefore needs to be refined.

According to the third observation in Section III-C, as the road regions extend away from the camera, they will be pro-

jected onto the higher portion of the image and will be associated with smaller disparity values. Mathematically, this can be formulated as

$$\Delta_{\text{roadprofile}(v)} \leq \Delta_{\text{roadprofile}(v+1)} \qquad (5)$$

where $\Delta_{\text{roadprofile}(v)}$ represents the disparity value of the road profile point in row v. Equation (5) is therefore used as a criterion to examine the validity of the initial road profile. This examination process is carried out from the bottom row to the highest row in the v-disparity image. If (5) is violated for some row, the next point with the largest intensity that conforms to (5) is determined as the new road profile point for that row.

There are two important issues that need to be addressed during the refinement of the road profile. First, (5) implies that the refinement of road profile for current row depends on its previous row. It is therefore very important to determine a suitable initial road profile point to begin the refinement process. As can be observed in the blue inset in Fig. 4(c), it is possible that the initial road profile point for the first row is not an actual road profile point when the disparity values for most of the pixels in the first row of the disparity map are uncertain or erroneous. At this time, it is an outlier. If we begin the refinement process from this row, the extraction of the following road profile points will be affected. From our experiments, we observe that a good road profile point to start the refinement process is one that possesses the largest disparity value among the first five rows. The row corresponding to that road profile point is treated as the starting row.

Second, (5) is able to distinguish the outliers from the initial road profile that appear on the right side of the actual road profile. However, in a few cases, outliers may appear on the left side of the actual road profile due to noisy input. As illustrated in Fig. 4(c), the outliers can be easily visually identified as they are located far from the previous rows. However, in the subsequent rows, the initial road profile points get closer again. If this situation is not corrected, (5) will cause the extracted road profile to deviate from the actual road profile in the subsequent rows during the refinement process. In order to resolve this issue, during the refinement of the initial road profile, if the current row's initial road profile point conforms to (5), the distance $\bar{\Delta} = \Delta_{\text{roadprofile}(v+1)} - \Delta_{\text{roadprofile}(v)}$ is checked. If $\bar{\Delta}$ is larger than some threshold value denoted as *threshold_close*, then the first road profile point in the subsequent rows within a limited range whose disparity value is larger than $\Delta_{\text{roadprofile}(v)}$ is found. The corresponding row is denoted as v'. If $\Delta_{\text{roadprofile}(v')} < \Delta_{\text{roadprofile}(v+1)}$, the initial road profile point for row v is then classified as an outlier. Then the point located in the range $[\Delta_{\text{roadprofile}(v+1-\text{threshold\_close})}, \Delta_{\text{roadprofile}(v+1)}]$ that has the largest intensity in the corresponding v-disparity image will be selected as the new road profile point.

C. Determination of the Horizon Line

According to the third observation in Section III-C, the further away the road is, the smaller the values v and Δ will become. Therefore, if the extracted road profile stops extending
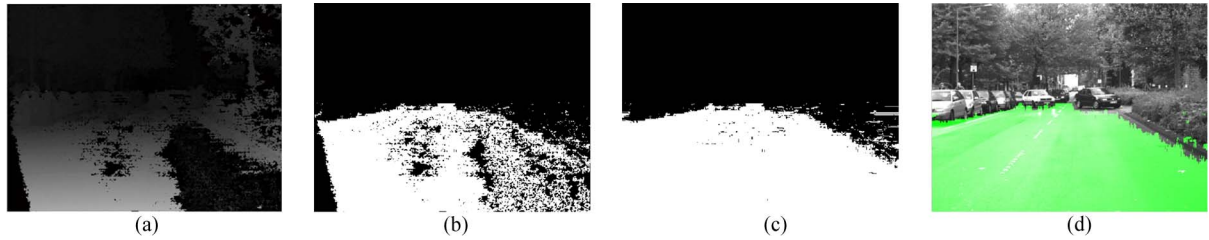
Fig. 5.   Illustration of road surface extraction process. (a) Noisy input disparity map with enhanced contrast for better visualization; (b) and (c) Intermediate results during the extraction process. (d) Final detected road surface.

toward the left for a predefined number (*threshold_largeobj*) of rows in the v-disparity image, this will mean that the road will cease to be visible in the road scene at that particular row. This row is the horizon line, as illustrated in Fig. 4(d). The part of extracted road profile under the horizon line is kept as the final road profile.

D. Road Surface Extraction

Once the longitudinal road profile is extracted, according to the fourth observation in Section III-C, it should be straightforward to extract the road surface as follows. For the image portion under the horizon line, the image points whose disparity values are smaller than or equal to the corresponding road profile are classified as road points; otherwise, they are regarded as obstacles. Regions with invalid disparity values are labeled according to their neighbors. This is the methodology adopted in [8]. However, the disparity map is often noisy. Some points in the same lateral road lines may not have the same disparity value in the given disparity map. For example, some road points may have larger disparity values than the majority of the road points for that scan line. Hence, this approach will result in many small blobs that are wrongly labeled.

We propose extracting the road surface in a more controlled fashion.

1) For each row v in the disparity map, the image point whose disparity value is smaller than or equal to $\Delta_{\mathrm{roadprofile}(v)}$ is classified as a road point. In addition, the image point whose disparity value is larger than $\Delta_{\mathrm{roadprofile}(v)}$ by certain degree denoted as *threshold_variance* and is not labeled as an obstacle point (based on the outcome of crude obstacles removal, as discussed in Section IV-A) is also classified as a road point.

2) The input disparity map may contain invalid regions whose disparity values are uncertain. Hence, it is necessary to perform interpolation for the invalid regions. The interpolation is performed in a scan line manner. For each continuous invalid span, we check its left and right neighboring spans to see if they are of the same label. If the left and right neighboring spans have the same label, then the invalid scan line is assigned the same label as its neighbors. Otherwise, the invalid scan line is assigned the same label as its larger neighbor.

3) At this time, there may be still some wrongly classified regions consisting of thin horizontal blobs within the correctly classified regions. Postprocessing must be then undertaken to filter out these wrongly classified blob

regions. The filtering process is carried out by checking its neighbors in a columnwise manner. For each column in the disparity map, we identify small vertical spans consisting of points with the same label. If the length of the span is smaller than *threshold_largeobj* and also notably smaller than the length of its neighboring spans, the points of the vertical span will be reassigned a new label. That is, if they are labeled as obstacles, they will be relabeled as roads and vice versa. The whole process of road surface extraction is illustrated in Fig. 5.

V. Experiments

A. Data Sets

We have evaluated our algorithm on three challenging benchmarks.

1) The enpeda data set [18] is a sequence of 394 640 ∗ 480 synthetic stereo frames containing both planar and nonplanar road scenarios. The ground truth disparity map is provided with subpixel accuracy for every pixel except for the occlusion part.

2) The KITTI benchmark [19] contains 194 1240 ∗ 376 stereo pairs and the corresponding semidense (approximately 50%) ground truth disparity maps. This benchmark mainly focuses on planar road scenarios. However, it covers quite an amount of different road contexts.

3) The Daimler benchmark [20] is a large-scale sequence of 21790 stereo frames captured in a busy urban environment with planar and nonplanar roads. As no ground truth disparity maps are provided for this benchmark, we have used the OpenCV implementation of the Semi-Global Matching algorithm [21] to generate the disparity maps. We use the default parameter settings in OpenCV. The obtained disparity maps are noisy and contain large invalid and erroneous regions.

The three chosen benchmarks constitute a comprehensive test bed that encompasses highly dynamic road situations. Some samples of the three benchmarks and the corresponding disparity maps are shown in Fig. 6.

B. Road Surface Detection Performance

Three parameters are required for the proposed algorithm, namely, *threshold_largeobj, threshold_close*, and *threshold_variance*, as introduced in Section IV. In this paper, *threshold_largeobj* is set to 12 for all three benchmarks. Since
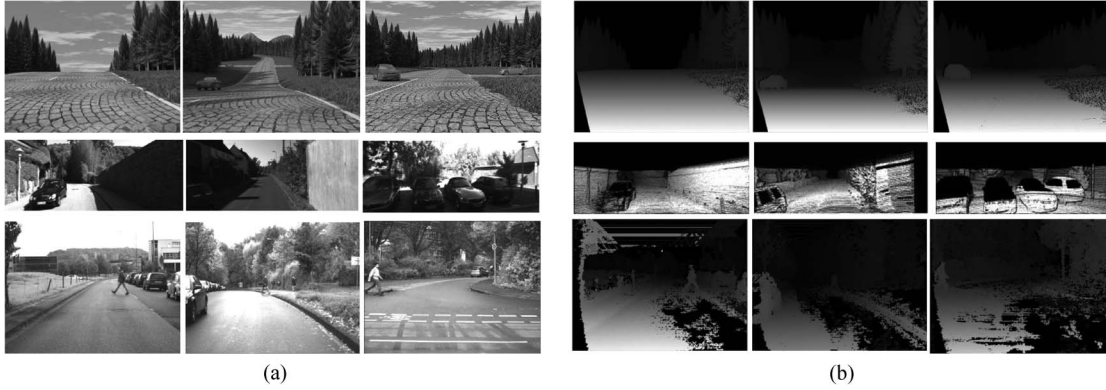
Fig. 6. Samples of (a) left images of the stereo pairs from (first row) the enpeda data set, (second row) the KITTI data set, and (third row) the Daimler data set and (b) their corresponding disparity maps for the three benchmarks used in our experiments. The contrast is enhanced for better visualization. The disparity maps serve as the inputs.
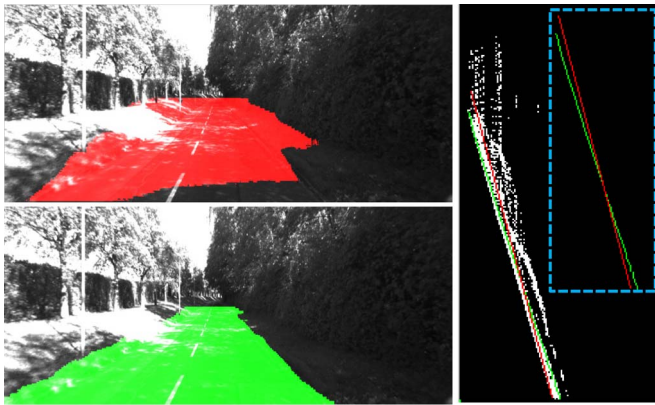

Fig. 7. Comparison for a planar road scenario. The top-left image shows the result of the baseline algorithm from [8], and the bottom-left image shows the result of the proposed algorithm. The right image shows the corresponding extracted road profiles of (red) the baseline algorithm and (green) the proposed algorithm. Blue inset highlights the two road profiles for better visualization.
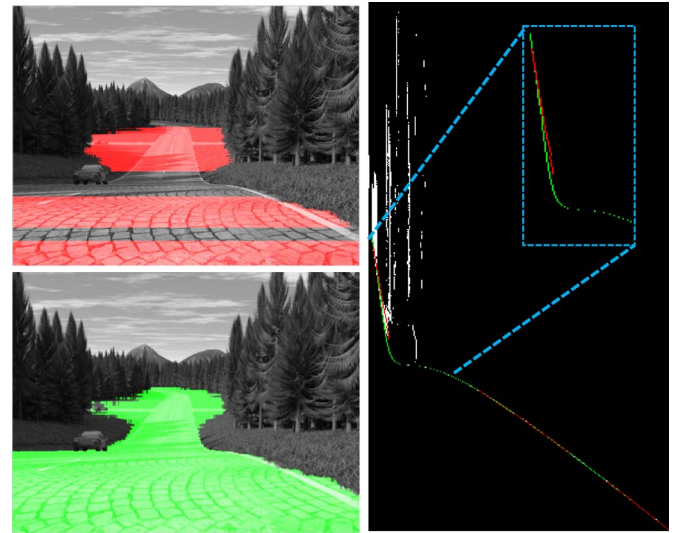

Fig. 8. Comparison for a nonplanar road scenario. The top-left image shows the result of the baseline algorithm from [8], and the bottom-left image shows the result of the proposed algorithm. The right image shows the corresponding extracted road profiles of (red) the baseline algorithm and (green) the proposed algorithm. Blue inset highlights a portion of the two road profiles for better visualization. Note that two vehicles are present in the image. The proposed algorithm can also correctly distinguish the vehicle that is further away, whereas the baseline algorithm fails.

subpixel accuracy is enabled and the ground truth disparity value is provided for every pixel in the enpeda data set, *threshold_close* is set to 2 and *threshold_variance* is set to 1 for this data set. For the other two data sets, *threshold_close* is set to 1 and *threshold_variance* is set to 2.

We have chosen the works in [8] and [13] as the baseline algorithms. The first baseline algorithm (i.e., [8]) is chosen as it is a representative work of the planar road assumption, which is the most widely used model to date. In addition, to the best of our knowledge, the work in [8] has presented the lowest computational complexity among all the reported stereo-based road surface detection methods in the literature. The second baseline algorithm (i.e., [13]) is one of the most recent works in the literature.

The work in [8] formulates the longitudinal road profile as a slanted straight line for planar road and a piecewise linear curve for nonplanar road in the v-disparity image. Hough transform is then utilized to extract the lines. The binary image input to the Hough transform is obtained by thresholding the v-disparity image with the value 25 in this experiment. For the planar road case, the longest line is extracted and treated as the road profile. Whereas, for the nonplanar road case, the ten highest Hough transform values are extracted. The upper or lower envelope of

these ten lines, depending on the accumulative gray value score, corresponds to the final road profile. For a fair comparison, all of these parameters are obtained after careful tuning. In addition, as explained in Section IV-D, the road surface extraction module in [8] is highly susceptible to noisy input. Again, in order to ensure a fair comparison, we have applied the proposed road surface extraction step to the baseline algorithm in order to increase the robustness of the baseline algorithm. Since the KITTI data set encompasses planar road scenarios, we enable the planar road assumption for the baseline algorithm on the KITTI benchmark. For the other two benchmarks, we enable the nonplanar road assumption for the baseline algorithm.

The main problem in [8] stems from the fact that it heavily relies on data fitting techniques, but it does not employ any countermeasure to deal with cases where the inputs to data fitting techniques are noisy. Not only roads but also obstacles will be projected as lines in the v-disparity. Hence, line
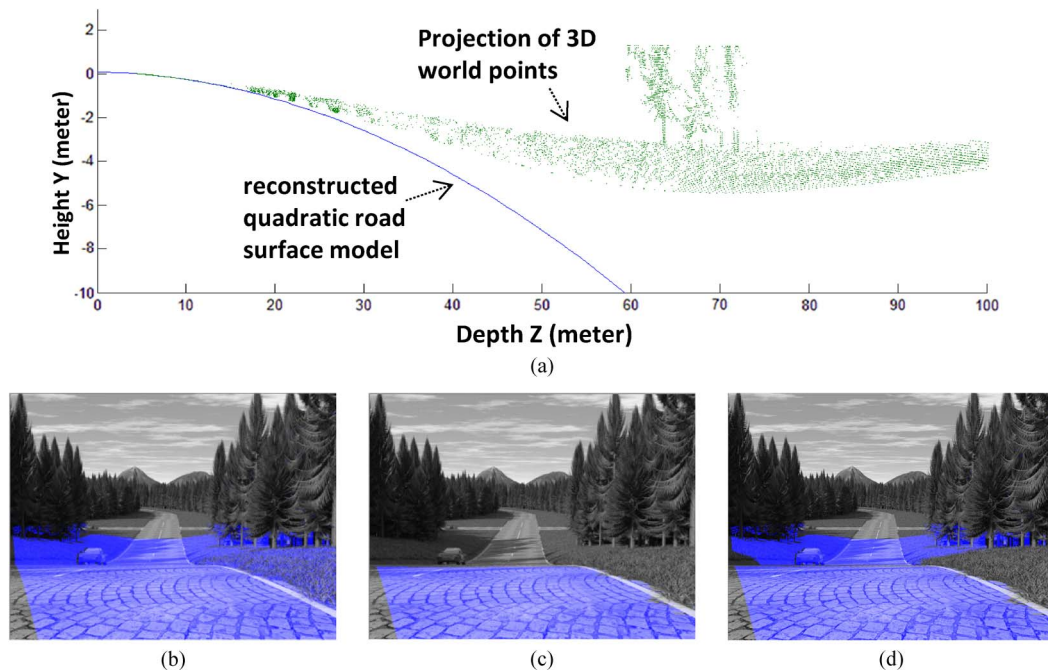
Fig. 9.   (a) Quadratic road model reconstructed by [13] only dovetails the realistic road surface in the vicinity of the vehicle. (b) Classification result of the density-based classifier from [13]. (c) Classification result of the road surface-based classifier from [13]. (d) Final classification result after fusing and filtering of (b) and (c).

extraction techniques like Hough transform will end up extracting multiple lines, and this makes it difficult to identify the line (planar) or a family of lines (nonplanar) that corresponds to the actual road profile. The advantage of our algorithm over [8] is clearly illustrated in Figs. 7 and 8. In Fig. 7, the two algorithms are compared for a planar road scenario. The effectiveness of using Hough transform to extract road profile is clearly impeded by the projection of the bushes along the side of the road. In Fig. 8, the two algorithms are compared using a nonplanar road scenario with an undulating hill. Since the baseline algorithm assumes that the road profile curvature is of constant sign, which is not true in this case, the baseline algorithm fails. It is noteworthy that the road profile extracted in Fig. 8 by the proposed method is not a regular curve that can be modeled mathematically. This clearly demonstrates that the proposed algorithm is not restricted to specific road scenarios but is able to accurately extract varying road profile shapes. In addition, it is evident from Fig. 8 that the proposed method is able to detect road at a large distance and correctly distinguish the vehicle that is very far away as a nonroad object.

Instead of working in the disparity space, [13] works in the 3-D digital elevation map (DEM) space. To distinguish road from nonroad entities, including obstacle and traffic isles, two classifiers are adopted. The density-based classifier marks DEM cells as road or obstacles based on the density of the reconstructed 3-D points. For the road surface-based classifier, the road surface is modeled such that quadratic variations of the height Y with the horizontal displacements X and depths Z are allowed. Then a combination of RANSAC, region growing, and least squares fitting is employed to compute the quadratic road surface. Based on the computed road surface model, road and nonroad entities are discriminated. Fusion and error filtering is finally performed on the results of the two classifiers. We have

adopted the parameter settings suggested by the authors in [13] for our experiments.

Compared with the planar road model, quadratic road model is more capable in some situations where the road surface presents quadratic curvature. However, due to its restricted parameterization, quadratic road model can only model slope changes in one direction. Hence, it will fail if the road is undulating. Fig. 9 presents the same world scenario as the one in Fig. 8 where the road surface is in an undulating shape. As shown in Fig. 9(a), the quadratic road model reconstructed by [13] only dovetails the realistic road surface in the vicinity of the vehicle. From the range of about 25 m, the reconstructed road model begins to fail.

Three-dimensional world points are reconstructed from the disparity space and reconstruction noise will be further introduced to the input data. As mentioned by the authors of [13], the reconstructed points' height uncertainty increases with depth and a 3-D road point at a depth of 30 m can have a height uncertainty of up to 17 cm. This makes the road surface-based classifier reliable only within a certain range. Due to this limitation, in the fusion and error filtering step, the work in [13] relies only on the results of the density-based classification for points at depth greater than 30 m.

The working principle for density-based classification is that different scene patches have different density values in the DEM. A double thresholding technique is designed by taking into account only the uphill road surface. Although the density-based classifier works fine with the planar and uphill road surface, it may fail for the case of downhill road surface. An obstacle standing on the downhill road may be misclassified as road depending on its height. As shown in Fig. 9(b), the vehicle on the downhill part of the road is wrongly classified by the density-based classifier. Fig. 9(c) shows the result of the road

Fig. 10.   Examples of detection results of the proposed algorithm for scenarios where the vicinity of the vehicle is filled with crowded objects.
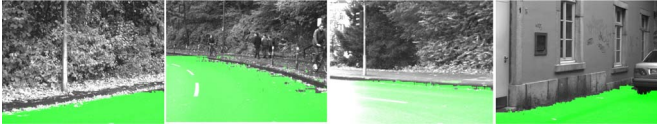

Fig. 11.   Examples of detection results of the proposed algorithm for scenarios where vehicle is turning. The corresponding yaw rate for the vehicle is 0.5256, −0.2144, −0.5625, and 0.3946. The unit of the yaw rate is in degrees per second.

surface based classifier. The final detection result is presented in Fig. 9(d), which is the fusion of Fig. 9(b) and (c). It can be observed that the algorithm proposed in [13] eventually wrongly classifies the vehicle as the road due to the intrinsic limitation of each classifier.

Crowded road scenarios are a kind of situation that needs to be given high attention as they are usually encountered in daily traffic. Since the proposed algorithm is well designed to remove the large obstacle at the first step, it is able to deal with this situation well. Some detection results from the proposed algorithm in crowded road scenarios are shown in Fig. 10.

Another special scenario is one where the vehicle is turning. The yaw rate depicts vehicle's turning degree. Although the yaw rate will not directly impact on our computations since the proposed algorithm does not utilize the temporal information, the high yaw rate may lead to changes in the roll angle. This may violate the assumption we have made in Section III-A that the cameras are installed such that the roll angle is negligible. Two factors help lessen vehicle's turning effect for the proposed algorithm. First, we are working on rectified images. The turning effect can be compensated by the image rectification algorithm to a large degree. Second, the proposed algorithm extracts the road surface in a controlled manner. Parameter *threshold_variance* helps further reduce the turning effect. The Daimler data set contains many cases where the vehicle is turning. Note that the data set has been rectified. Some examples of the detection results from the proposed algorithm for scenarios where vehicle is turning at a high yaw rate are presented in Fig. 11. Our evaluations show that the proposed algorithm also works fine in these scenarios.

Finally, we carry out both qualitative and quantitative evaluations between the proposed and the baseline algorithms. In the following, the work in [8] is denoted as *Baseline_A*, and the work in [13] is denoted as *Baseline_B*. Note that the data sets used in this evaluation contain many invalid regions whose disparity values are not available. For the image points within these regions, the corresponding 3-D coordinates are not reconstructed and therefore cannot be classified. To ensure a fair comparison, the interpolation technique for the invalid regions proposed in Section IV-D has been also applied to *Baseline_B*. In addition to Figs. 7–9, more qualitative evaluation results are presented in Figs. 12 and 13, respectively. Note that the testing
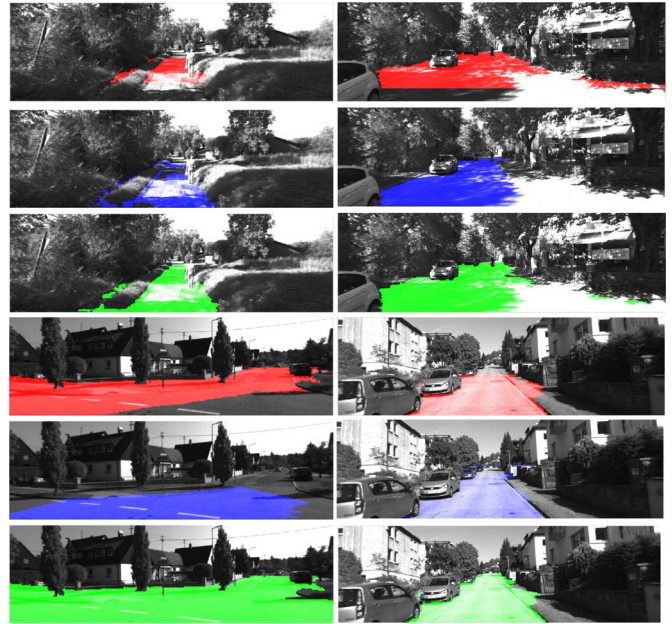

Fig. 12.   More comparisons between (red) *Baseline_A*, (blue) *Baseline_B*, and (green) proposed algorithm for KITTI data set.
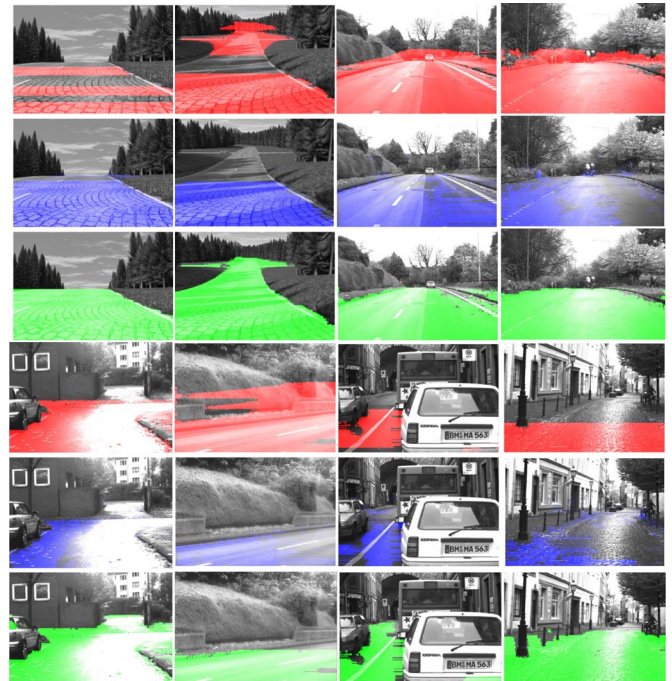

Fig. 13.   More comparisons between (red) *Baseline_A*, (blue) *Baseline_B*, and (green) proposed algorithm for enpeda and Daimler data sets. The images within the first two columns in the higher part are from enpeda data set, whereas the others are from Daimler data set.

samples in Figs. 12 and 13 include different scenarios such as planar road scenarios; uphill, downhill, and undulating hill nonplanar road scenarios; the scenario where the vicinity of the vehicle is filled with crowded objects; and the scenario where the vehicle is turning. As explained earlier, *Baseline_A* is frail in these complex scenarios. *Baseline_B* works well with the KITTI data set since the KITTI data set mainly contains planar or uphill road scenarios. In addition, the disparity maps used in this data set are ground truth; hence, the reconstructed 3-D

TABLE I
CONTINGENCY TABLE

		Ground Truth	
		Non-Road	Road
Result	Non-Road	TN	FN
	Road	FP	TP

TABLE II
FOUR PIXELWISE METRICS FOR PERFORMANCE EVALUATION

Pixel-wise Metric	Definition
Quality	Q = TP/TP+FP+FN
Detection Rate	P = TP/TP+FP
Detection Accuracy	R = TP/TP+FN
Effectiveness	F = 2PR/P+R

TABLE III
COMPARISONS OF THE DETECTION PERFORMANCE BETWEEN
THE BASELINE AND PROPOSED ALGORITHMS

		Q	P	R	F
enpeda	*Baseline_A*	64.17%	93.66%	67.08%	78.17%
	Baseline_B	81.33%	98.70%	82.21%	89.71%
	Proposed	88.34%	95.62%	92.07%	93.81%
	Improvement_A	24.17%	1.96%	24.99%	15.64%
	Improvement_B	7.01%	-3.08%	9.86%	4.10%
KITTI	*Baseline_A*	79.25%	84.92%	92.23%	88.42%
	Baseline_B	78.24%	93.15%	83.02%	87.79%
	Proposed	82.09%	86.32%	94.36%	90.16%
	Improvement_A	2.84%	1.40%	2.13%	1.74%
	Improvement_B	3.85%	-6.83%	11.34%	2.37%
Daimler	*Baseline_A*	80.72%	90.40%	88.29%	89.33%
	Baseline_B	63.22%	89.32%	68.39%	77.46%
	Proposed	85.16%	92.47%	91.51%	91.99%
	Improvement_A	4.44%	2.07%	3.22%	2.66%
	Improvement_B	21.94%	3.15%	23.12%	14.53%

points present high accuracy. It is noteworthy that the traffic isles in this data set are also correctly discriminated from the road by *Baseline_B*, as shown in Fig. 12. Since enpeda data set contains many frames where the road surface is undulating, *Baseline_B* can only detect the road surface correctly in the vicinity of the vehicle for these frames. Daimler data set encompasses all the dynamic road shapes, including planar, uphill, downhill, and undulating hill road scenarios. In addition, the corresponding disparity maps are quite noisy and contain large invalid and erroneous regions. The data noisy will be further amplified during the process of 3-D reconstruction from the disparity space. In order to distinguish the road surface from traffic isles, the classification threshold for road in *Baseline_B* is set conservatively. These three factors contribute to large amount of false detection in *Baseline_B*. The visual comparison in Figs. 12 and 13 clearly demonstrates the superiority of the proposed algorithm over the baseline algorithms in various challenging road scenarios.

For the quantitative evaluation, we adopt the evaluation framework in [7]. Manual road labeling is performed on all the left images in the enpeda, KITTI data sets, and a subset of the Daimler data set containing 1613 frames to generate the ground truth. The Daimler data set consists of a 27-min sequence of video. We have chosen the frames within the 10th and 20th minute periods of the video sequence for manual labeling. The 10th minute of the Daimler sequence depicts a situation where the vehicle is moving on a planar road with crowded obstacles in the vicinity of the vehicle. The 20th minute Daimler sequence is a situation that encompasses all the dynamic road situations such as planar, uphill, downhill, and undulating hill road scenarios. In addition, many cases where the vehicle is turning are present in this subsequence. Based on the ground truth and the detection results, each pixel in the test samples is labeled as one of the four cases: TP, TN, FP, and FN according to Table I. Then four metrics are adopted to describe the detection performance: *Quality* (Q), *Detection rate* (P), *Detection accuracy* (R), and *Effectiveness* (F) as formulated in Table II. Each of the metrics provides a different insight. For the detailed interpretation of these metrics, please refer to [7].

The quantitative evaluation results have been summarized in Table III. In Table III, we can observe that the proposed algorithm outperforms *Baseline_A* for all the metrics on all the test cases. The proposed algorithm also achieves better results than *Baseline_B* for all the metrics except the detection rate on

enpeda and KITTI data sets. The proposed algorithm obtains a slightly lower detection rate than *Baseline_B* on these two data sets due to the following reasons: For enpeda data set, the proposed method misclassifies the grass field at a distance of more than 150 m as road surface. Whereas, for the KITTI data set, some of the traffic isles whose height is quite close to road surface are misclassified by the proposed algorithm as road surface. The quantitative evaluation results are consistent with the qualitatively results.

It is noteworthy that, although the input disparity maps for the KITTI data set only achieve semidense density (approximately 50%) and the input disparity maps for the Daimler data set are noisy and contain large invalid and erroneous regions, the proposed algorithm can still achieve high detection performance. This clearly demonstrates that the proposed algorithm is robust to noisy input. For the enpeda data set, the major contributing factor that prevents the detection performance of the proposed algorithm from achieving 100% performance lies in the existence of the grass field present from the 71th frame to the 200th frame in the data set. An example of this is illustrated in Fig. 8. As shown, the grass field extends far away from the camera (more than 150 m) and the difference between the heights of the grass field and the road surface is small. Therefore, the observed disparity values for the grass field and the disparity values for the road surface nearby are almost the same. Due to this, the proposed algorithm classifies these grass fields as the road surface.

C. Runtime Performance

The proposed algorithm does not require complex calculations (i.e., only integer addition and comparison operations are needed.) The runtime bottleneck of the work in [8] mainly lies in the Hough transform employed to extract the lines in the v-disparity image.

The work in [13] presents much higher computational complexity. Dense 3-D points need to be reconstructed. Canny edge detector and RANSAC-based plane fitting are employed for initial surface fitting. The biggest bottleneck for [13] lies in the uncertainty model-driven surface growing period. During this period, the quadratic surface is recomputed, in a least squares

TABLE IV
RUNTIME COMPARISON BETWEEN THE BASELINE
AND PROPOSED APPROACHES

		enpeda	KITTI	Daimler
Baseline_A (Seconds)	Total	4.423	2.640	236.168
	Per Frame	0.011	0.014	0.011
Baseline_B (Seconds)	Total	41.429	41.121	2187.59
	Per Frame	0.105	0.212	0.100
Proposed (Seconds)	Total	2.749	2.054	131.09
	Per Frame	0.007	0.011	0.006
Speedup_A(%)		37.85	22.20	44.50
Speedup_B (%)		93.36	95.00	94.01

fashion, about N times per frame on average. The value of N depends on the road surface types, the quality and number of the input points, and the selection of cells for initial surface fitting. N can be up to 200, as reported in [13]. For our experiments conducted on the three adopted data sets, we observe that N can be up to 300. Many intensive computations with double precision data type are required.

To compare the runtime performance, we have implemented the proposed algorithm and both of the baseline algorithms in C++ on an Intel Core 2 3.16-GHz machine. In order to make a fair comparison, we have employed the highly optimized implementation of Hough transform in OpenCV for the implementation of [8] and Canny edge detector for [13]. In addition, for our implementation of [13], the expected road density maps are computed offline for each data set and not included into the runtime measurement. During the stage of region growing, the matrix used to compute the quadratic surface model defined by (5) in [13] is updated using only the newly added DEM cells for each iteration. Compared with the experimental setup in [13], the image size is larger in our experiments. In addition, the 3-D points are reconstructed in software instead of using specific hardware.

The results in Table IV show that the proposed algorithm easily achieves real-time performance (less than 0.015 s/frame). Therefore, the proposed algorithm is highly suitable for in-vehicle deployments, which often have limited computation resources. In addition, the proposed algorithm is about 35% faster than the baseline algorithm in [8] and about 94.13% faster than the work in [13] on average. Note that the runtime does not include the computation time of the disparity map for both the baseline algorithms and the proposed algorithm.

VI. CONCLUSION

Road surface detection is usually required as an initial step in many applications (e.g., autonomous navigation, object detection, and tracking) to provide the geometrical constraint to facilitate the subsequent step. In this paper, we have proposed a simple but efficient nonparametric depth based road surface detection algorithm that is inspired by four intrinsic road attributes observed under stereo geometry. Unlike existing methods that rely on rigid mathematical models, the proposed nonparametric algorithm has been shown to be capable of tackling highly dynamic road scenarios. This has paved the way for overcoming the limitations of existing parametric methods that cannot cope

with cases where the road profile does not fit the predefined model or when the constantly varying road profiles cannot be modeled mathematically. Extensive experimental results using three challenging benchmarks (i.e., enpeda, KITTI, and Daimler) show that the proposed algorithm outperforms the baseline algorithms both in terms of detection accuracy and runtime performance. The data set used to evaluate the proposed technique includes different scenarios such as various planar road scenarios; uphill, downhill, and undulating hill nonplanar road scenarios; the scenarios where the vicinity of the vehicle is filled with crowded objects; and scenarios involving turning vehicles. The experimental results show that the proposed algorithm achieves both high detection and runtime performance, and hence, it is well suited for deployment in a wide range of applications involving advanced driver assistance.

REFERENCES

[1] A. Bar Hillel, R. Lerner, D. Levi, and G. Raz, "Recent progress in road and lane detection: A survey," *Mach. Vis. Appl.*, vol. 25, no. 3, pp. 727–745, Apr. 2014.

[2] C. Thorpe, M. H. Hebert, T. Kanade, and S. A. Shafer, "Vision and navigation for the Carnegie-Mellon Navlab," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 10, no. 3, pp. 362–373, May 1988.

[3] B. Southall and C. J. Taylor, "Stochastic road shape estimation," in *Proc. IEEE ICCV*, 2001, vol. 1, pp. 205–212.

[4] H. Yinghua, W. Hong, and Z. Bo, "Color-based road detection in urban traffic scenes," *IEEE Trans. Intell. Transp. Syst.*, vol. 5, no. 4, pp. 309–318, Dec. 2004.

[5] C. Rasmussen, "Grouping dominant orientations for ill-structured road following," in *Proc. IEEE CVPR*, 2004, vol. 1, pp. I-470–I-477.

[6] K. Hui, J. Y. Audibert, and J. Ponce, "Vanishing point detection for road detection," in *Proc. IEEE CVPR*, 2009, pp. 96–103.

[7] J. M. Alvarez, T. Gevers, and A. M. Lopez, "3D Scene priors for road detection," in *Proc. IEEE CVPR*, 2010, pp. 57–64.

[8] R. Labayrade, D. Aubert, and J. P. Tarel, "Real time obstacle detection in stereovision on non flat road geometry through 'v-disparity' representation," in *Proc. IEEE Intell. Veh. Symp.*, 2002, vol. 2, pp. 646–651.

[9] D. Thao and C. Hoffmann, "Fast object hypotheses generation using 3D position and 3D motion," in *Proc. IEEE CVPRW*, 2005, p. 56.

[10] S. J. Krotosky and M. M. Trivedi, "On color-, infrared-, and multimodal-stereo approaches to pedestrian detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 4, pp. 619–629, Dec. 2007.

[11] A. D. Sappa, F. Dornaika, D. Ponsa, D. Geronimo, and A. Lopez, "An efficient approach to onboard stereo vision system pose estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 3, pp. 476–490, Sep. 2008.

[12] N. Soquet, D. Aubert, and N. Hautiere, "Road segmentation supervised by an extended V-disparity algorithm for autonomous navigation," in *Proc. IEEE Intell. Veh. Symp.*, 2007, pp. 160–165.

[13] F. Oniga and S. Nedevschi, "Processing dense stereo data using elevation maps: Road surface, traffic isle, and obstacle detection," *IEEE Trans. Veh. Technol.*, vol. 59, no. 3, pp. 1172–1182, Mar. 2010.

[14] S. Nedevschi *et al.*, "High accuracy stereovision approach for obstacle detection on non-planar roads," in *Proc. IEEE Intell. Eng. Syst.*, 2004, pp. 1–6.

[15] A. Wedel *et al.*, "B-spline modeling of road surfaces with an application to free-space estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 10, no. 4, pp. 572–583, Dec. 2009.

[16] C. G. Keller *et al.*, "The benefits of dense stereo for pedestrian detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1096–1106, Dec. 2011.

[17] A. Broggi, C. Caraffi, P. P. Porta, and P. Zani, "The single frame stereo vision system for reliable obstacle detection used during the 2005 DARPA grand challenge on TerraMax," in *Proc. IEEE Intell. Transp. Syst. Conf.*, 2006, pp. 745–752.

[18] T. Vaudrey, C. Rabe, R. Klette, and J. Milburn, "Differences between stereo and motion behaviour on synthetic and real-world stereo sequences," *Proc. Image Vis. Comput. New Zealand*, pp. 1–6, 2008.

[19] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE CVPR*, 2012, pp. 3354–3361.

[20] C. G. Keller, M. Enzweiler, and D. M. Gavrila, "A new benchmark for stereo-based pedestrian detection," in *Proc. IEEE IV Symp.*, 2011, pp. 691–696.

[21] H. Hirschmuller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.

Siew-Kei Lam (M'03) received the B.A.Sc. (Honors) degree and the M.Eng. degree in computer engineering from Nanyang Technological University (NTU), Singapore.

Since 1994 he has been with NTU, where he is currently a Research Fellow with the Centre for High Performance Embedded Systems and has worked on a number of challenging projects that involved porting of complex algorithms in very large-scale integration. He is also familiar with rapid prototyping and application-specific integrated-circuit design flow methodologies. His research interests include embedded system design algorithms and methodologies, algorithms-to-architectural translations, and high-speed arithmetic units.

Meiqing Wu received the M.S. degree in computer engineering from Peking University, Beijing, China, in 2009. She is currently working toward the Ph.D. degree in the Centre for High Performance Embedded Systems, Nanyang Technological University, Singapore.

Her research interests include stereo vision, motion analysis, object detection, and tracking for urban traffic scene understanding.

Thambipillai Srikanthan (SM'92) received the B.Sc. degree in computer and control systems and the Ph.D. degree in system modelling and information systems engineering from Coventry University, Coventry, U.K.

He served in academia in the U.K. for eight years before joining Nanyang Technological University (NTU), Singapore, in 1991. Currently, he holds a Full Professor and joint appointments as the Director of a 100 strong Centre for High Performance Embedded Systems (CHiPES) and the Director of the Intelligent Devices and Systems (IDeAS) Cluster at NTU. He founded CHiPES in 1998 and elevated it to a university-level research center in February 2000. He has published more than 250 technical papers. His research interests include design methodologies for complex embedded systems, architectural translations of compute intensive algorithms, computer arithmetic, and high-speed techniques for image processing and dynamic routing.