Road Surface 3D Reconstruction Based on Dense Subpixel Disparity Map Estimation

Rui Fan, *Graduate Student Member, IEEE*, Xiao Ai, and Naim Dahnoun

***Abstract*— Various 3D reconstruction methods have enabled civil engineers to detect damage on a road surface. To achieve the millimeter accuracy required for road condition assessment, a disparity map with subpixel resolution needs to be used. However, none of the existing stereo matching algorithms are specially suitable for the reconstruction of the road surface. Hence in this paper, we propose a novel dense subpixel disparity estimation algorithm with high computational efficiency and robustness. This is achieved by first transforming the perspective view of the target frame into the reference view, which not only increases the accuracy of the block matching for the road surface but also improves the processing speed. The disparities are then estimated iteratively using our previously published algorithm, where the search range is propagated from three estimated neighboring disparities. Since the search range is obtained from the previous iteration, errors may occur when the propagated search range is not sufficient. Therefore, a correlation maxima verification is performed to rectify this issue, and the subpixel resolution is achieved by conducting a parabola interpolation enhancement. Furthermore, a novel disparity global refinement approach developed from the Markov random fields and fast bilateral stereo is introduced to further improve the accuracy of the estimated disparity map, where disparities are updated iteratively by minimizing the energy function that is related to their interpolated correlation polynomials. The algorithm is implemented in C language with a near real-time performance. The experimental results illustrate that the absolute error of the reconstruction varies from 0.1 to 3 mm.**

***Index Terms*— 3D reconstruction, road condition assessment, subpixel disparity estimation, parabola interpolation, Markov random fields, fast bilateral stereo.**

I. Introduction

THE condition assessment of asphalt and concrete civil infrastructures, e.g., bridges, tunnels and pavements, is essential to ensure their usability while still providing maximum safety for the users. It also allows the government to allocate the limited resources for maintenance and appraise long-term investment schemes [1]. The manual visual inspections performed by either structural engineers or certified inspectors are cost-intensive, time-consuming and cumbersome [2]. In 2014, a one-off investment of £12bn was suggested by the Asphalt Industry Alliance to improve the road condition across England and Wales [3]. Over the last decade, various technologies such as remote sensing, vibration sensing and computer vision have been increasingly applied in civil engineering to assess the physical and functional condition of the infrastructures such as potholes, cracking, etc.

The remote sensing methods which have been used in satellites, airplanes, unmanned aerial vehicles or multi-purpose survey vehicles have indeed reduced the workload of inspectors. However, the traditional geotechnical methods can never be entirely replaced by the remote sensing approaches [4]. Using accelerometers and GPS for data acquisition, vibration-based methods always cause distress misdetection in spite of their advantages of small storage requirements, cost-effectiveness and real-time performance [2]. As for the approaches based on 2D computer vision, the spatial structure of the road surface cannot be illustrated explicitly [2]. Therefore, 3D reconstruction-based methods are more feasible to overcome these disadvantages and simultaneously provide an enhancement in terms of detection accuracy and processing efficiency.

3D reconstruction methods can be classified as laser scanner-based, Microsoft Kinect-based and passive sensor-based. The laser scanner collects the reflected laser pulse from an object to construct its accurate 3D model [4]. Although it provides accurate modeling results, the laser scanner equipment used for road condition analysis is still costly [2]. As for the methods based on the Microsoft Kinect sensor, the depth measurement for the outdoor environment is somewhat ineffective, especially for materials which strongly absorb the infrared light [5]. Therefore, the passive sensor-based methods, e.g., stereo vision, are more capable of reconstructing the 3D road surface for condition assessment or damage detection.

To reconstruct a real-world environment with passive sensing techniques, multiple camera views are required [6]. Images from different viewpoints can be captured using either a single movable camera or an array of cameras [7]. In this paper, we use a ZED stereo camera to acquire a pair of images for road surface 3D reconstruction. Since the stereo rig is assumed to be well-calibrated, the main work performed in this paper is the disparity estimation. The algorithms for disparity estimation can be classified as local, global and semi-global. Local algorithms simply match a series of blocks and select the correspondence with the lowest cost or the

Manuscript received July 18, 2017; revised December 30, 2017 and February 12, 2018; accepted February 17, 2018. Date of publication February 22, 2018; date of current version March 27, 2018. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Patrick Le Callet. *(Corresponding author: Rui Fan.)*

R. Fan is with the Visual Information Group, University of Bristol, Bristol BS8 1UB, U.K. (e-mail: ranger_fan@outlook.com).

X. Ai is with the Quantum Technology Enterprise Centre, Nanoscience and Quantum Information Building, University of Bristol, Bristol BS8 1FD, U.K. (e-mail: xiao.ai@bristol.ac.uk).

N. Dahnoun is with the Department of Electrical and Electronic Engineering, Merchant Venturers Building, University of Bristol, Bristol BS8 1UB, U.K. (e-mail: naim.dahnoun@bristol.ac.uk).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the author. The material includes a demo video. The total size of the video is 64.4 MB. Contact ranger_fan@outlook.com for further questions about this work.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIP.2018.2808770

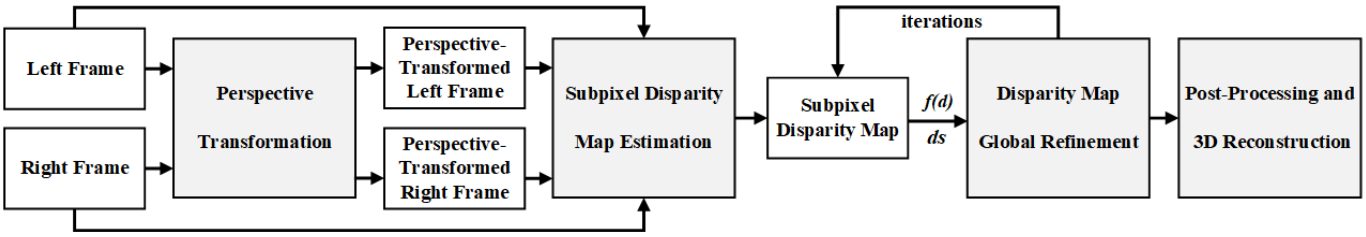


Fig. 1. Stereo vision-based road surface 3D reconstruction system workflow.

highest correlation. This optimization is also known as winner-take-all (WTA). Unlike local algorithms, global algorithms process the stereo matching using some more sophisticated optimization techniques, e.g., Graph Cut (GC) [8] and Belief Propagation (BP) [9]. These algorithms are commonly developed based on the Markov Random Fields (MRF) [10], where finding the best disparities is formulated as a probability maximization problem. This is later addressed by energy minimization approaches. Semi-global matching (SGM) [11] approximates the MRF inference by performing cost aggregation along all directions in the image and this greatly improves the accuracy and efficiency of stereo matching. However, the occlusion problem always makes it difficult to find the optimum value for the smoothness parameters: over-penalizing the smoothness term can help avoid the ambiguities around discontinuities but on the other hand can lead to errors for continuous areas [12]. Therefore, some authors have proposed to break down the global problem into multiple local problems, each of which is affected by uncertainties to a lesser extent [13]. For instance, one alternative way of setting smoothness parameters is to group pixels in the image into different slanted planes [13]–[15]. Disparities in different plane groups are estimated with local constraints. However, this results in high computational complexities, making real-time performance challenging.

In order to further improve the trade-off between speed and accuracy, seed-and-grow local algorithms have been used extensively. In these algorithms, the disparity map is grown from a selection of seeds to minimize expensive computations and reduce mismatches caused by ambiguities. For example, Sara [16], [17] and Cech and Sara [18] presented an efficient quasi-dense stereo matching algorithm, named growing correspondence seeds (GCS), to estimate disparities iteratively with the search range propagated from a collection of reliable seeds. Similarly, various Delaunay triangulation-based stereo matching algorithms (DTSM) have been proposed in [19]–[21] to estimate tunable semi-dense disparity maps with the support of a piecewise planar mesh. Our previous algorithm [22], [23] also provides an efficient strategy for local stereo matching whereby the search range on row v is propagated from three estimated neighboring disparities on row $v+1$. Our algorithm performs better than GCS and DTSM in terms of estimating dense disparity maps for road scenes where the road disparities decrease gradually from the bottom to the top, while the disparities of obstacles remain the same. The aim of this paper is to reconstruct the road scenes for pothole detection. In this regard, the proposed disparity estimation algorithm is developed based on our previous work in [23]. To assess the condition of a road surface, millimeter accuracy is desired in 3D reconstruction and thus disparities in subpixel resolution are inevitable. Therefore, the correlation costs around the initial disparity are interpolated into a parabola and the position of the extrema is selected as the subpixel disparity.

However, the subpixel disparity maps obtained from parabola interpolation are still unsatisfactory because the correlation costs of neighborhood systems are not aggregated before finding the best disparities. To aggregate neighboring costs adaptively, some authors have proposed to filter the whole cost volume with a bilateral filter since it provides a feasible solution for the initial message passing problem on a fully connected MRF [12]. These algorithms are also known as Fast Bilateral Stereo (FBS) [24]–[26]. However, the intensive computational complexity introduced when filtering the whole cost volume severely impact on the processing speed. In this regard, we believe that only the candidates around the best disparities need to be processed and a novel disparity refinement approach is proposed in this work. The workflow of our stereo vision-based road surface 3D reconstruction system is depicted in Fig. 1.

Firstly, the perspective view of the road surface in the target image is transformed into its reference view, which greatly enhances the similarity of the road surface between the two images. Since the propagated search range is sometimes insufficient, the desirable disparities have to be further verified to ensure they possess the highest correlation costs. The latter ensures the feasibility of parabola interpolation-based subpixel enhancement. To further optimize the obtained subpixel disparity map, the interpolated parabola functions $f(d)$ are set as the labels in the MRF because they contain the information of both disparity values and correlation costs. By updating the parabola functions $f(d)$ and subpixel disparities d_s iteratively, a disparity in a continuous area becomes smooth but it is preserved when discontinuities occur. Finally, each 3D point on the road surface is computed based on its projections on the left and right images. The reconstruction accuracy is evaluated using three sample models (see section VI-A for more details). Our datasets are publicly available at: http://www.ruirangerfan.com.

The rest of the paper is structured as follows: section II presents a novel perspective transformation (PT) method. In section III, we describe a subpixel disparity estimation algorithm. A disparity map global refinement approach is introduced in section IV. In section V, the disparity map is post-processed and the 3D road surface is reconstructed.

In section VI, the experimental results are illustrated and the performance of the proposed algorithm is evaluated. Finally, section VII summarizes the paper and provides some recommendations for future work.

II. Perspective Transformation

In this paper, the proposed algorithm focuses entirely on the road surface which can be treated as a ground plane (GP). To enhance the accuracy of stereo matching, we first draw on the concept of ground plane constraint in [27] and [28] to transform the perspective views of two images before estimating their disparities. GP constraint is commonly used in a wide range of obstacle detection systems, where the image on one side is set as the reference and the other image is transformed into the reference view. Pixels arising from the GP satisfy the same affine transformation while an object above the GP will not be transformed successfully [27]. Referring to the experimental results in [28], pixels from an obstacle are distorted in the transformed image. Nevertheless, the GP in the transformed image looks more similar to its reference view. Therefore, a perspective transformation makes the obstacle areas noisy and unreliable but greatly enhances the similarity of the road surface between two images. In this paper, the road surface is defined as:

$$\boldsymbol{n}^\top \boldsymbol{P_w} + \beta = 0 \tag{1}$$

where $\boldsymbol{P_w} = [X_w, Y_w, Z_w]^\top$ is an arbitrary 3D point on the road surface. Its projections on the left image π_l and the right image π_r are $\boldsymbol{p_l} = [u_l, v_l]^\top$ and $\boldsymbol{p_r} = [u_r, v_r]^\top$, respectively. $\boldsymbol{n} = [n_0, n_1, n_2]^\top$ is the normal vector of the road surface. The planar transformation between $\tilde{\boldsymbol{p}}_{\boldsymbol{l}} = [u_l, v_l, 1]^\top$ and $\tilde{\boldsymbol{p}}_{\boldsymbol{r}} = [u_r, v_r, 1]^\top$ is given in Eq. 2 [6]. Here, $\tilde{\boldsymbol{p}} = [u, v, 1]^\top$ denotes the homogeneous coordinate of $\boldsymbol{p} = [u, v]^\top$.

$$\tilde{\boldsymbol{p}}_{\boldsymbol{r}} = \boldsymbol{H_{rl}} \tilde{\boldsymbol{p}}_{\boldsymbol{l}} \tag{2}$$

$\boldsymbol{H_{rl}} \in \mathbb{R}^{3\times3}$ denotes a homograph matrix, which is generally used to distinguish obstacles from the road surface [27]. It can be decomposed as [6]:

$$\boldsymbol{H_{rl}} = \boldsymbol{K_r} \left(\boldsymbol{R_{rl}} - \frac{\boldsymbol{T_{rl}} \boldsymbol{n}^\top}{\beta} \right) \boldsymbol{K_l}^{-1} \tag{3}$$

where $\boldsymbol{R_{rl}}$ is a SO(3) matrix and $\boldsymbol{T_{rl}}$ is a translation vector. $\boldsymbol{P_l}$ in the left camera coordinate system can be transformed to $\boldsymbol{P_r}$ in the right camera coordinate system according to $\boldsymbol{P_r} = \boldsymbol{R_{rl}} \boldsymbol{P_l} + \boldsymbol{T_{rl}}$. $\boldsymbol{K_l}$ and $\boldsymbol{K_r}$ are intrinsic matrices of the two cameras. For a well-calibrated stereo system, $\boldsymbol{R_{rl}}$, $\boldsymbol{T_{rl}}$, $\boldsymbol{K_l}$ and $\boldsymbol{K_r}$ are already known. We only need to estimate $\boldsymbol{n}$ and β for $\boldsymbol{H_{rl}}$. Generally, $\boldsymbol{H_{rl}}$ can be estimated with at least four pairs of correspondences $\boldsymbol{p_l}$ and $\boldsymbol{p_r}$ [6]. Hattori et al. proposed a pseudo-projective camera model where several assumptions are made about road geometry to simplify the estimation of $\boldsymbol{H_{rl}}$ [27]. In this paper, we improve on their algorithm by considering the following hypotheses:

- $\boldsymbol{K_l}$ and $\boldsymbol{K_r}$ are identical.
- $\boldsymbol{R_{rl}}$ is an identity matrix.
- $\boldsymbol{T_{rl}}$ is in the same direction as the X_w-axis.
- the road surface is a horizontal plane: $n_1 Y_w + \beta = 0$.

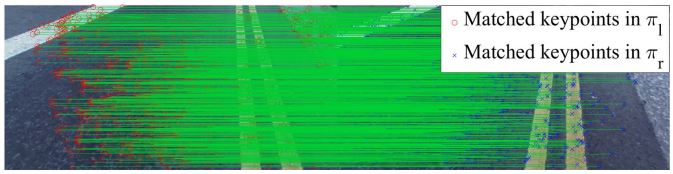


Fig. 2. BRISK-based on-road keypoints detection and matching between the left and right images.

Algorithm 1 Perspective Transformation

Data: π_l and π_r
Result: $\boldsymbol{\alpha} = [\alpha_0, \alpha_1]^\top$
detect and match the keypoints in π_l and π_r;
if $|v_{li} - v_{ri}| > \epsilon$ *or* $u_{li} - u_{ri} < 0$ **then**
  remove $\boldsymbol{p_{li}}$ and $\boldsymbol{p_{ri}}$ from $\boldsymbol{Q_l}$ and $\boldsymbol{Q_r}$, respectively;
estimate $\boldsymbol{\alpha}$ using the least squares fitting;
all points in the target image are shifted $\alpha_0 + \alpha_1 v - \delta$ pixels to the reference view;

- rotation of the stereo rig is only about the X_w-axis.

For a perfectly-calibrated stereo rig, $v_l = v_r = v$. The disparity is defined as $d = u_l - u_r$. The projection of a horizontal plane on the v-disparity map is a linear pattern [29]:

$$d = -\frac{T_c n_1}{\beta}(f \sin\theta - v_0 \cos\theta) - v\frac{T_c n_1}{\beta}\cos\theta = \alpha_0 + \alpha_1 v \tag{4}$$

where θ is the pitch angle between the stereo rig and the road surface (an example can be seen in Fig. 7 (a)), f is the focus length of the cameras, T_c is the baseline, and (u_0, v_0) is the principal point in pixels. When $\theta = \pi/2$, $d = -f T_c n_1/\beta$ is a constant. Otherwise, d is proportional to v [29]. This implies that a perspective distortion always exists for the GP in two images, which further affects the accuracy of block matching. Therefore, the PT aims to make the GP in the transformed image similar to that in the reference frame.

Now, the PT can be straightforwardly realized using parameters $\boldsymbol{\alpha} = [\alpha_0, \alpha_1]^\top$. The proposed PT is detailed in algorithm 1. $\boldsymbol{\alpha}$ can be estimated by solving a least squares problem with a set of reliable correspondences $\boldsymbol{Q_l} = [\boldsymbol{p_{l1}}, \boldsymbol{p_{l2}}, \ldots, \boldsymbol{p_{lm}}]^\top$ and $\boldsymbol{Q_r} = [\boldsymbol{p_{r1}}, \boldsymbol{p_{r2}}, \ldots, \boldsymbol{p_{rm}}]^\top$. In this paper, we use BRISK (Binary Robust Invariant Scalable Keypoints) to detect and match $\boldsymbol{Q_l}$ and $\boldsymbol{Q_r}$. It allows a faster execution to achieve approximately the same number of correspondences as SIFT (Scale-Invariant Feature Transform) and SURF (Speeded-Up Robust Features) [30]. An example of on-road keypoints detection and matching is illustrated in Fig. 2.

Since outliers can severely affect the accuracy of least squares fitting, we first remove the less reliable correspondences before estimating $\boldsymbol{\alpha}$, where ϵ is proposed to be 1. For the left disparity map ℓ^{lf} estimation, each point on row v in π_r is shifted $\alpha_0 + \alpha_1 v - \delta$ pixels to the right, where δ is a constant set to 20 (for dataset 1 and 2) or 30 (for dataset 3) to guarantee that all the disparities are positive. Similarly, each point in π_l is shifted $\alpha_0 + \alpha_1 v - \delta$ pixels to the left when π_r is served

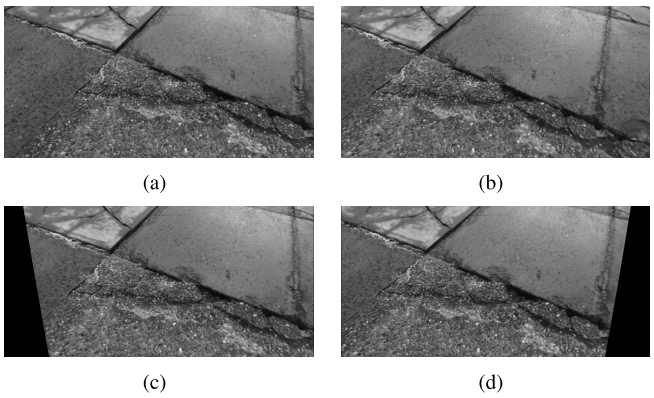


Fig. 3. Perspective transformation. (a) left image. (b) right image. (c) transformed right image. (d) transformed left image. (a) and (c) are used as the input left and right images for the left disparity map estimation. (d) and (b) are used as the input left and right images for the right disparity map estimation.

as the reference. An example of perspective transformation is presented in Fig. 3. The performance improvements achieved by using the PT will be discussed in section VI.

III. Subpixel Disparity Map Estimation

As compared to many other stereo matching algorithms which aim at automotive applications, the trade-off between speed and precision has been greatly improved in our previous work [22], [23]. The subpixel accuracy can be achieved by conducting a parabola interpolation for the correlation costs around the initial disparity [24]. The subpixel disparity global refinement will be discussed in section IV.

A. Stereo Matching

In this paper, our previous algorithm [22] is utilized to estimate integer disparities, where the NCC (Normalized Cross-Correlation) is used to compute the matching costs, and the search range SR for pixel at (u, v) is propagated from three estimated neighboring disparities on row $v + 1$. To accelerate the NCC execution, we rearrange the NCC equation as follows:

$$c(u,v,d)=\frac{1}{n\sigma_l\sigma_r}\left(\sum_{x=u-\rho}^{x=u+\rho}\sum_{y=v-\rho}^{y=v+\rho} i_l(x,y)i_r(x-d,y)-n\mu_l\mu_r\right) \tag{5}$$

where $c(u, v, d)$ is defined as the correlation cost between two square blocks selected from π_l and π_r, and a higher $c(u, v, d)$ corresponds to a better matching and vice-versa. i_l or i_r is the intensity of a pixel in π_l or π_r. The edge length of the square block is $2\rho + 1$, and n represents the number of pixels in it. (u, v) and $(u-d, v)$ are the centers of the left and right blocks, respectively. μ_l and μ_r denote the means of the intensities within the two blocks. σ_l and σ_r are their standard deviations.

From Eq. 5, μ and σ only matter for each independent block selected from π_l or π_r, and d determines a pair of blocks for matching. Therefore, the calculation of μ_l, μ_r, σ_l and σ_r will always be repeated in conventional NCC-based stereo matching algorithms. In [23], we propose to pre-calculate the values of μ and σ and store them in a static program storage for direct indexing. Thus, the computational complexity of the NCC is simplified to a dot product, making stereo matching more efficient. More details on the implementation procedure are available in [23].

Algorithm 2 Correlation Maxima Verification

Data: disparity map ℓ
Result: correlation maxima verified disparity map ℓ_{cmv}
if $c(u, v, d) > \max\{c(u, v, d-1), c(u, v, d+1)\}$ **then**
  $\ell_{cmv}(u, v) \leftarrow \ell(u, v)$;
else if $c(u, v, d-1) < c(u, v, d) < c(u, v, d+1)$ **then**
  repeat
    compute $c(u, v, d+k)$, $k \geq 2$;
  until $c(u, v, d+k) < c(u, v, d+k-1)$;
  $\ell_{cmv}(u, v) \leftarrow d + k - 1$;
else
  repeat
    compute $c(u, v, d-k)$, $k \geq 2$;
  until $c(u, v, d-k) < c(u, v, d-k+1)$;
  $\ell_{cmv}(u, v) \leftarrow d - k + 1$;
end

1) Search Range Propagation (SRP): Since the concept of "local coherence constraint" was proposed in [31], many researchers have turned their focus on seed-and-grow algorithms for stereo matching. Either semi-dense or quasi-dense disparity maps can be estimated efficiently with the guidance from a collection of reliable feature points [16]–[21]. In this paper, the road surface is treated as a GP whose disparities change gradually from the bottom of the image to its top, which makes our previous algorithm [22] more efficient than other methods in terms of estimating an accurate dense disparity map. The proposed algorithm propagates the search range SR iteratively row by row from the bottom of the image to its top. In the first iteration, the disparity estimation performs a full search range. Then, SR at (u, v) is propagated from three estimated neighboring disparities using Eq. 6, where τ is the bound of SR and is set as 1 in this paper. The left and right disparity maps, ℓ^{lf} and ℓ^{rt}, are shown in Fig. 4 (a) and (b), respectively.

$$SR = \bigcup_{k=u-1}^{u+1} \{sr | sr \in [\ell(k, v+1) - \tau, \ell(k, v+1) + \tau]\} \tag{6}$$

2) Correlation Maxima Verification (CMV): Since the search range propagates using Eq. 6, errors may occur in subpixel enhancement when $c(u, v, d-1)$ or $c(u, v, d+1)$ is not computed and compared with $c(u, v, d)$. Therefore, CMV will run until the correlation cost of the disparity is a local maxima. More details are provided in algorithm 2.

B. Left-Right Consistency (LRC) Check

Due to the fact that each pair of correspondences from two images is unique, if we select an arbitrary pixel (u, v) from the left disparity map ℓ^{lf}, there should exist at most one correspondence in the right disparity map ℓ^{rt} [12]:

$$\ell^{lf}(u, v) = \ell^{rt}(u - \ell^{lf}(u, v), v) \tag{7}$$

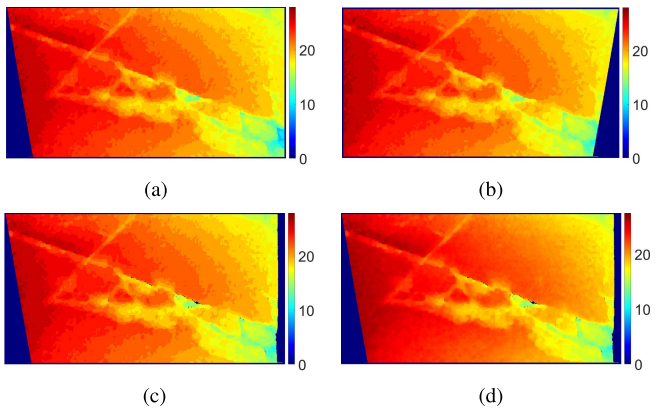


Fig. 4. Subpixel disparity map estimation. (a) left disparity map. (b) right disparity map. (c) left disparity map processed with the LRC check. (d) subpixel disparity map.

Pixels that are only visible in one disparity map are marked as uncertainties. A LRC check is performed to remove these half-occluded areas. Although the LRC check doubles the computational complexity by re-projecting the estimated disparities from one disparity map to the other one, most of the infeasible conjugate pairs can be removed, and an outlier in the disparity map can be found. The left disparity map after the LRC check processing is illustrated in Fig. 4 (c).

C. Subpixel Enhancement

In this paper, the road surface application requires a millimeter accuracy in 3D reconstruction. A disparity error larger than one pixel may result in a non-neglected difference in the reconstructed road surface [32]. Therefore, subpixel resolution is inevitable to achieve a highly accurate result.

For each pixel whose disparity d is $\ell(u, v)$, we fit a parabola to three correlation costs $c(u, v, d-1)$, $c(u, v, d)$ and $c(u, v, d+1)$ around the initial disparity d. The centerline of the parabola is selected as the subpixel displacement d_s as follows [26]:

$$d_s = d + \frac{c(u,v,d-1) - c(u,v,d+1)}{2c(u,v,d-1) + 2c(u,v,d+1) - 4c(u,v,d)} \tag{8}$$

Since the CMV guarantees that $c(u, v, d)$ is larger than both $c(u, v, d-1)$ and $c(u, v, d+1)$, d_s will be between $d-1$ and $d+1$. Fig. 4 (c) after the subpixel enhancement is given in Fig. 4 (d).

IV. Disparity Map Global Refinement

A. Markov Random Fields and Fast Bilateral Stereo

Unlike the principle of WTA applied in local stereo matching algorithms, the matching costs from neighboring pixels are also taken into account in global algorithms, e.g., GC and BP. The MRF is a commonly used graphical model in these global algorithms. An example of the MRF model is depicted in Fig. 5.

The graph $\mathcal{G} = (\mathcal{P}, \mathcal{E})$ is a set of vertices $\mathcal{P}$ connected by edges $\mathcal{E}$, where $\mathcal{P} = \{\boldsymbol{p_{11}}, \boldsymbol{p_{12}}, \cdots, \boldsymbol{p_{mn}}\}$ and $\mathcal{E} = \{(\boldsymbol{p_{ij}}, \boldsymbol{p_{st}}) \mid \boldsymbol{p_{ij}}, \boldsymbol{p_{st}} \in \mathcal{P}\}$. Two edges sharing one common vertex are called a pair of adjacent edges [33]. Since the MRF is considered to be undirected, $(\boldsymbol{p_{ij}}, \boldsymbol{p_{st}})$ and $(\boldsymbol{p_{st}}, \boldsymbol{p_{ij}})$ refer to the same edge here. $\mathcal{N}_{ij} = \{\boldsymbol{n_{1p_{ij}}}, \boldsymbol{n_{2p_{ij}}}, \cdots, \boldsymbol{n_{kp_{ij}}} \mid \boldsymbol{n_{p_{ij}}} \in \mathcal{P}\}$ is a neighborhood system for $\boldsymbol{p_{ij}}$.

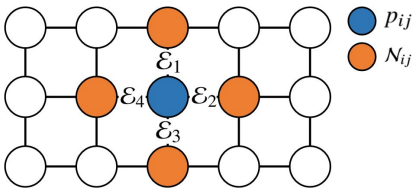


Fig. 5. Markov random fields.

For stereo vision problems, $\mathcal{P}$ is a $m \times n$ disparity map and $\boldsymbol{p_{ij}}$ is a vertex (or node) at the site of (i, j) with a label of disparity d_{ij}. Because more candidates taken into consideration usually make the inference of a true disparity intractable, only the neighbors adjacent to $\boldsymbol{p_{ij}}$ are considered for stereo matching [10]. This is also known as a pairwise MRF. In this paper, $k = 4$ and $\mathcal{N}$ is a four-connected neighborhood system. $\mathcal{E}_1 = (\boldsymbol{p_{ij}}, \boldsymbol{n_{1p_{ij}}})$, $\mathcal{E}_2 = (\boldsymbol{p_{ij}}, \boldsymbol{n_{2p_{ij}}})$, $\mathcal{E}_3 = (\boldsymbol{p_{ij}}, \boldsymbol{n_{3p_{ij}}})$ and $\mathcal{E}_4 = (\boldsymbol{p_{ij}}, \boldsymbol{n_{4p_{ij}}})$ are adjacent edges sharing the vertex $\boldsymbol{p_{ij}}$. The disparity of $\boldsymbol{p_{ij}}$ tends to have a strong correlation with its vicinities, while it is linked implicitly to any other random nodes in the disparity map. In [10], the joint probability of the MRF is written as:

$$P(\boldsymbol{p}, q) = \prod_{\boldsymbol{p_{ij}} \in \mathcal{P}} \Phi(\boldsymbol{p_{ij}}, q_{\boldsymbol{p_{ij}}}) \prod_{\boldsymbol{n_{p_{ij}}} \in \mathcal{N}_{ij}} \Psi(\boldsymbol{p_{ij}}, \boldsymbol{n_{p_{ij}}}) \tag{9}$$

where $q_{\boldsymbol{p_{ij}}}$ represents the intensity differences, $\Phi(\cdot)$ expresses the compatibility between possible disparities and the corresponding intensity differences, and $\Psi(\cdot)$ expresses the compatibility between $\boldsymbol{p_{ij}}$ and its neighborhood system. Now, the aim of finding the best disparity is equivalent to maximizing the probability in Eq. 9. This can be realized by formulating Eq. 9 as an energy function [10]:

$$E(\boldsymbol{p}) = \sum_{\boldsymbol{p_{ij}} \in \mathcal{P}} D(\boldsymbol{p_{ij}}, q_{\boldsymbol{p_{ij}}}) + \sum_{\boldsymbol{n_{p_{ij}}} \in \mathcal{N}_{ij}} V(\boldsymbol{p_{ij}}, \boldsymbol{n_{p_{ij}}}) \tag{10}$$

$D(\cdot)$ and $V(\cdot)$ are two energy functions. $D(\cdot)$ corresponds to the matching cost and $V(\cdot)$ determines the aggregation from the neighbors. In the MRF model, the method to formulate an adaptive $V(\cdot)$ is important because the intensity in discontinuous areas usually varies greatly from that of its neighbors [34]. Since Tomasi et al. introduced the bilateral filter in [35], many authors have investigated its applications to aggregate the matching costs [24]–[26]. These methods are also grouped into fast bilateral stereo, where both intensity difference and spatial distance provide a weight to adaptively constrain the aggregation of discontinuities. A general representation of the cost aggregation in FBS is represented as follows:

$$c_{agg}(i, j, d) = \frac{\sum_{x=i-\rho}^{i+\rho} \sum_{y=j-\rho}^{j+\rho} \omega_d(x, y)\omega_r(x, y)c(x, y, d)}{\sum_{x=i-\rho}^{i+\rho} \sum_{y=j-\rho}^{j+\rho} \omega_d(x, y)\omega_r(x, y)} \tag{11}$$

where ω_d is based on the spatial distance and ω_r is based upon the color similarity. The costs c within a square block are aggregated adaptively to obtain c_{agg}.

Although the FBS has shown a good performance in terms of matching accuracy, it usually takes a long time to process the whole cost volume. Therefore, we propose an improved adaptive aggregation method to optimize the subpixel disparity map iteratively.

B. Subpixel Disparity Refinement With Energy Minimization

In this paper, the local algorithm proposed in section III greatly minimizes the trade-off between accuracy and speed. A precise subpixel disparity map can be estimated with a near real-time performance. Compared to conventional MRF-based algorithms, our global refinement method only aggregates the costs around the best disparity and updates the disparity map in a more efficient way. The proposed disparity refinement algorithm is developed based on the following assumptions:

- the subpixel disparity map obtained in section III is acceptable.
- for an arbitrary pixel, its neighbors (excluding discontinuities) in all directions have similar disparities.
- the interpolated parabola $f(d) = \beta_0 + \beta_1\ d + \beta_2\ d^2$ in section III-C is locally smooth.

Before going into further details about our disparity refinement approach, we first rewrite the energy function in Eq. 10 in a more general way as follows [36]:

$$E(\boldsymbol{p}) = E_{data}(\boldsymbol{p_{ij}}) + \lambda E_{smooth}(\boldsymbol{p_{ij}}, \boldsymbol{n}_{p_{ij}}) \tag{12}$$

where the term E_{data} penalizes the solutions that are inconsistent with the observed data, E_{smooth} enforces the piecewise smoothness and λ is the smoothness parameter. For conventional MRF-based stereo matching algorithms, E_{data} denotes the matching cost and E_{smooth} is the cost aggregation from the neighborhood system. By minimizing the global energy of the whole random field, a disparity map can be estimated.

In section III-C, we fit a parabola $f(d) = \beta_0 + \beta_1\ d + \beta_2\ d^2$ to three correlation costs $c(u, v, d-1)$, $c(u, v, d)$ and $c(u, v, d+1)$ to get the subpixel disparity d_s. The parabola function $f(d)$ contains the information of both subpixel disparity and correlation costs. Since $f(d)$ is assumed to be locally smooth, the neighboring pixels tend to have similar parabola parameters. However, when an abrupt change occurs, they vary significantly and in this case, the condition for uniform smoothness is no longer valid. Therefore, we use function $f(d_{p_{ij}})$ as the label in MRF. By adaptively aggregating functions $f(d_{\boldsymbol{n}_{p_{ij}}})$ of the neighborhood system to $f(d_{p_{ij}})$, $f(d_{p_{ij}})$ is updated iteratively.

In order to ensure energy minimization rather than energy maximization as widely presented in literature, the term E_{data} is defined as:

$$E_{data}(\boldsymbol{p_{ij}}) = -f(d_{\boldsymbol{p_{ij}}}) \tag{13}$$

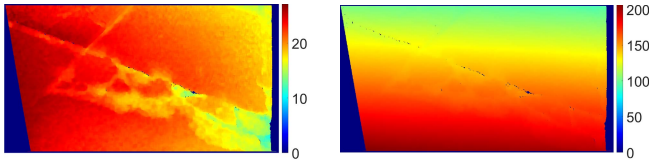
λ has a value of $1/\sqrt{2}$ in this paper. Using the same strategy of adaptive aggregation in FBS, we define the smoothness energy $E_{smooth}(\boldsymbol{p_{ij}}, \boldsymbol{n}_{p_{ij}})$ as the adaptive sum of negative interpolated parabolas $-f(d_{\boldsymbol{n}_{p_{ij}}})$ of spatially varying horizontal and vertical nearest neighbors:

$$E_{smooth}(\boldsymbol{p_{ij}}, \boldsymbol{n}_{p_{ij}}) = -\sum_{m=1}^{k} \omega(\boldsymbol{p_{ij}}, \boldsymbol{n_{mp_{ij}}}) f(d_{\boldsymbol{n_{mp_{ij}}}}) \tag{14}$$

where

$$\omega(\boldsymbol{p_{ij}}, \boldsymbol{n_{mp_{ij}}}) = \exp\left\{-\frac{||\mathcal{E}_m||_2^2}{{\sigma_d}^2}\right\} \exp\left\{-\frac{(d_{\boldsymbol{n_{mp_{ij}}}} - d_{\boldsymbol{p_{ij}}})^2}{{\sigma_r}^2}\right\} \tag{15}$$

The weighting coefficient ω is determined by both the spatial distance $||\mathcal{E}_m||_2$ between $\boldsymbol{n_{mp_{ij}}}$ and $\boldsymbol{p_{ij}}$ and the difference between $d_{\boldsymbol{n_{mp_{ij}}}}$ and $d_{\boldsymbol{p_{ij}}}$. σ_d and σ_r are two parameters used to control ω and they are respectively set to 1 and 5 in this paper. If $d_{\boldsymbol{n_{mp_{ij}}}}$ is similar to $d_{\boldsymbol{p_{ij}}}$, the weight for cost aggregation is higher. The energy function with respect to the correlation costs is updated iteratively. The subpixel disparity map is optimized by approximating the minima of the updated energy functions. In this paper, the proposed process is iterated three times, and the result after the third iteration is shown in Fig. 6 (a).



(a) (b)

Fig. 6. Disparity map global refinement and post-processing. (a) subpixel disparity map after the third iteration. (b) post-processed disparity map.

V. Post-Processing and 3D Reconstruction

Due to the fact that the perspective views have been transformed in section II, the estimated subpixel disparities on row v should be added $\alpha_0 + \alpha_1 v - \delta$ to obtain the post-processed disparity map which is illustrated in Fig. 6 (b). Then, the intrinsic and extrinsic parameters of the stereo system are used to compute each 3D point $\boldsymbol{P}_w = [X_w, Y_w, Z_w]^\top$ from its projections $\boldsymbol{p_l} = [u_l, v_l]^\top$ and $\boldsymbol{p_r} = [u_r, v_r]^\top$, where v_r is equivalent to v_l, and u_r is associated with u_l by disparity d.

For many state-of-the-art road model estimation algorithms, the effects caused by the non-zero roll angle (Fig. 7 (b)) are always ignored because the stereo cameras will not change significantly over time [37]. However, the experimental set-up in this paper is installed manually and the roll angle may introduce a distortion on the v-disparity histogram. Therefore, the roll angle needs to be estimated for the initial frame to minimize its impact on the perspective transformation for the rest of the sequences. As in [37], the roll angle γ can be estimated by fitting a linear plane $(d(u, v) = \gamma_0 + \gamma_1 u + \gamma_2 v)$ to a small patch from the near field in the disparity map and $\gamma = \arctan(-\gamma_1/\gamma_2)$. The pitch angle θ can be estimated by rearranging Eq. 4 as Eq. 16, where the parameters $[\alpha_0, \alpha_1]^\top$ have been approximated in section II. The yaw angle ψ shown

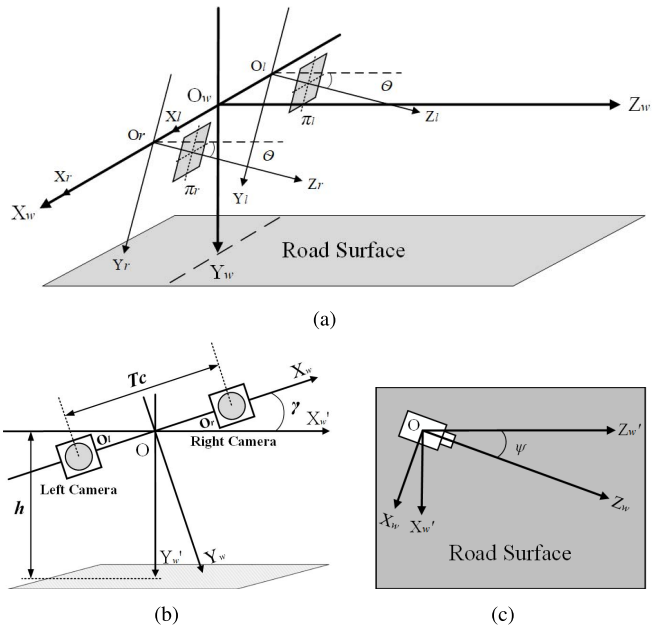


Fig. 7. Extrinsic rotations. (a) pitch angle θ. (b) roll angle γ. (c) yaw angle ψ. h is the height of the proposed binocular system.

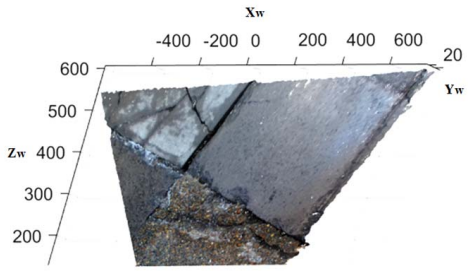


Fig. 8. Road surface 3D reconstruction.

in Fig. 7 (c) is assumed to be 0.

$$\theta = \arctan\left(\frac{1}{f}\left(\frac{\alpha_0}{\alpha_1} + v_0\right)\right) \tag{16}$$

Each 3D point $[X_w, Y_w, Z_w]^\top$ can be transformed into $[X'_w, Y'_w, Z'_w]^\top$ using Eq. 17 [38]. The rotation matrix $\boldsymbol{R} = \boldsymbol{R}_\psi \boldsymbol{R}_{\boldsymbol{\theta}} \boldsymbol{R}_{\boldsymbol{\gamma}}$ is a SO(3) matrix. The rotation with $\boldsymbol{R}$ makes pothole detection much easier. The 3D reconstruction of Fig. 3 (a) is illustrated in Fig. 8.

$$\begin{bmatrix} X'_w \\ Y'_w \\ Z'_w \end{bmatrix} = \boldsymbol{R}_\psi \boldsymbol{R}_{\boldsymbol{\theta}} \boldsymbol{R}_{\boldsymbol{\gamma}} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} \tag{17}$$

where

$$\boldsymbol{R}_\psi = \begin{bmatrix} \cos\psi & 0 & \sin\psi \\ 0 & 1 & 0 \\ -\sin\psi & 0 & \cos\psi \end{bmatrix} \tag{18}$$

$$\boldsymbol{R}_{\boldsymbol{\theta}} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix} \tag{19}$$

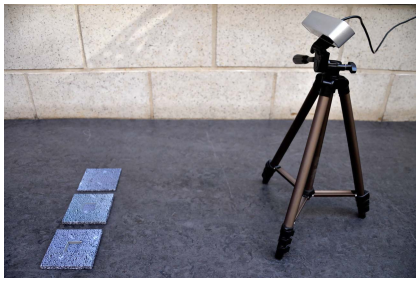


Fig. 9. Experimental set-up.

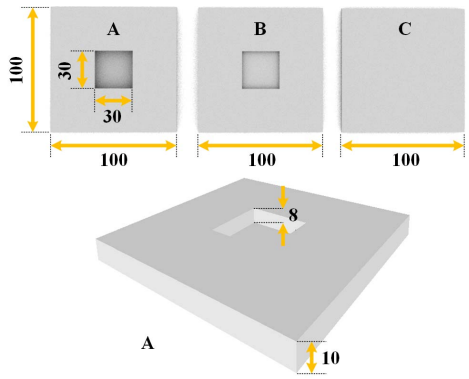


Fig. 10. Designed 3D sample models. The unit is millimeter.

$$\boldsymbol{R}_{\boldsymbol{\gamma}} = \begin{bmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{20}$$

VI. Experimental Results

In this section, we evaluate the performance of our proposed road surface 3D reconstruction algorithm both qualitatively and quantitatively. The algorithm is programmed in C language on an Intel Core i7-4720HQ CPU using a single thread. The following subsections detail the experimental set-up and the performance evaluation.

A. Experimental Set-Up

In our experiments, a state-of-the-art stereo camera from ZED Stereolabs is used to capture 1080p (3840×1080) videos at 30 fps or 2.2K (4416×1242) videos at 15 fps [39]. The baseline is 120 mm. With its ultra sharp six element all-glass dual lenses and 16:9 native sensors, the video is 110° wide-angle and able to cover the scene up to 20 m. An example of the experimental set-up is shown in Fig. 9. The stereo camera is calibrated manually using the stereo calibration toolbox from MATLAB R2017a. The overall calibration mean error in pixels is 0.335.

To quantify the accuracy of the proposed algorithm, we designed three sample models A, B and C with different sizes. They are printed with a MakerBot Replicator 2 Desktop 3D Printer whose layer resolution is from 0.1 mm to 0.3 mm. Their top views and the stereogram of model A are illustrated in Fig. 10, where A and B are designed with grooves to

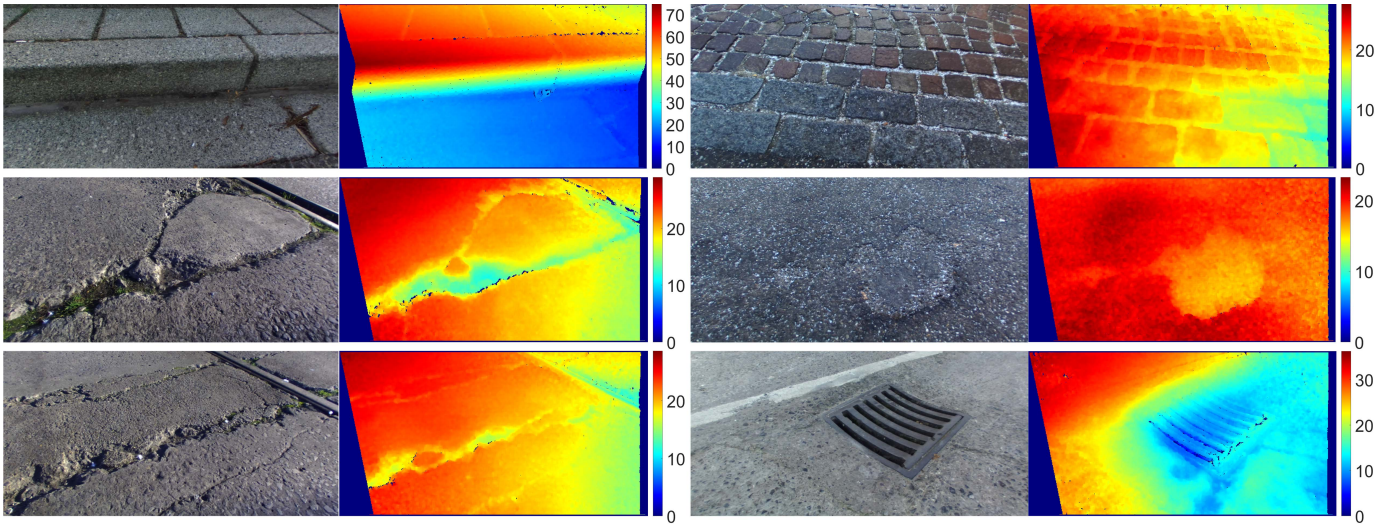


Fig. 11. Experimental results. The first and third columns are the input left images. The second and fourth columns are the subpixel disparity map without post-processing.

TABLE I
DESIGN SIZE AND ACTUAL SIZE OF THE SAMPLE MODELS

Sample model	Design size (mm×mm×mm)		Actual size (mm×mm×mm)	
	Model	Groove	Model	Groove
A	$100.00 \times 100.00 \times 10.00$	$30.00 \times 30.00 \times 8.00$	$99.97 \times 99.83 \times 10.31$	$29.74 \times 30.01 \times 8.25$
B	$100.00 \times 100.00 \times 10.00$	$30.00 \times 30.00 \times 3.00$	$100.39 \times 100.10 \times 9.82$	$30.28 \times 29.98 \times 3.52$
C	$100.00 \times 100.00 \times 5.00$	n/a	$100.00 \times 99.98 \times 5.92$	n/a

simulate potholes. To get the ground truth for our experiments, we measured the actual size of these models using an electronic vernier caliper. Both the design and actual sizes of the models are presented in Table I. Since the models are printed with a single color, resulting in homogeneous areas, we attached them with a piece of paper with the texture of the road surface printed on it to avoid the ambiguities during stereo matching, as can be seen in Fig. 9.

Using the above experimental set-up, we create three datasets (91 stereo image pairs) for the road surface 3D reconstruction. Datasets 1 and 2 aim at road sceneries, and dataset 3 contains the sample models to help researchers qualify their reconstruction results. The datasets are available at: http://www.ruirangerfan.com.

The following subsections analyze the performance of our algorithm in terms of disparity accuracy, reconstruction accuracy and processing speed.

B. Disparity Evaluation

Some examples of the disparity maps are illustrated in Fig. 11. Before estimating the disparity map, we transform the target image into its reference view, which greatly eliminates the perspective distortion for a GP between two images. Since the GP in the left and right images now looks similar to each other, the average of the highest correlation costs goes higher, which is depicted in Fig. 12. For stereo matching with only SRP, the average of the highest correlation increases gradually from 0.807 ($\rho = 1$) to 0.845 ($\rho = 4$). However, when ρ goes above 4, c keeps decreasing. If we

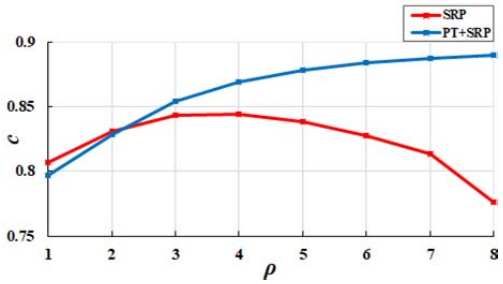


Fig. 12. Comparison between SRP and PT+SRP in terms of the average of the highest correlation costs.

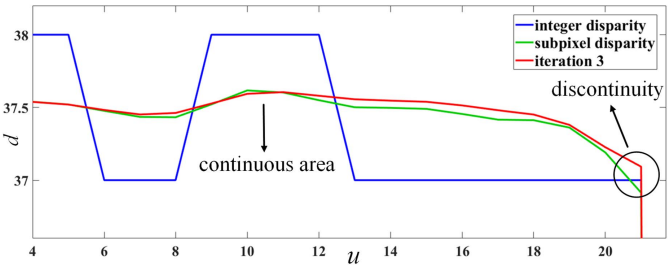


Fig. 13. Evaluation of subpixel enhancement and disparity global refinement.

pre-process the input image pairs with the PT, the average of the highest correlation costs in the SRP stereo will grow gradually between $\rho = 1$ and $\rho = 8$. In this paper, our datasets are created with high-resolution images, and ρ is proposed to be 5. Compared with the conventional SRP stereo, the PT improves the average correlation cost with an increase of 0.05.

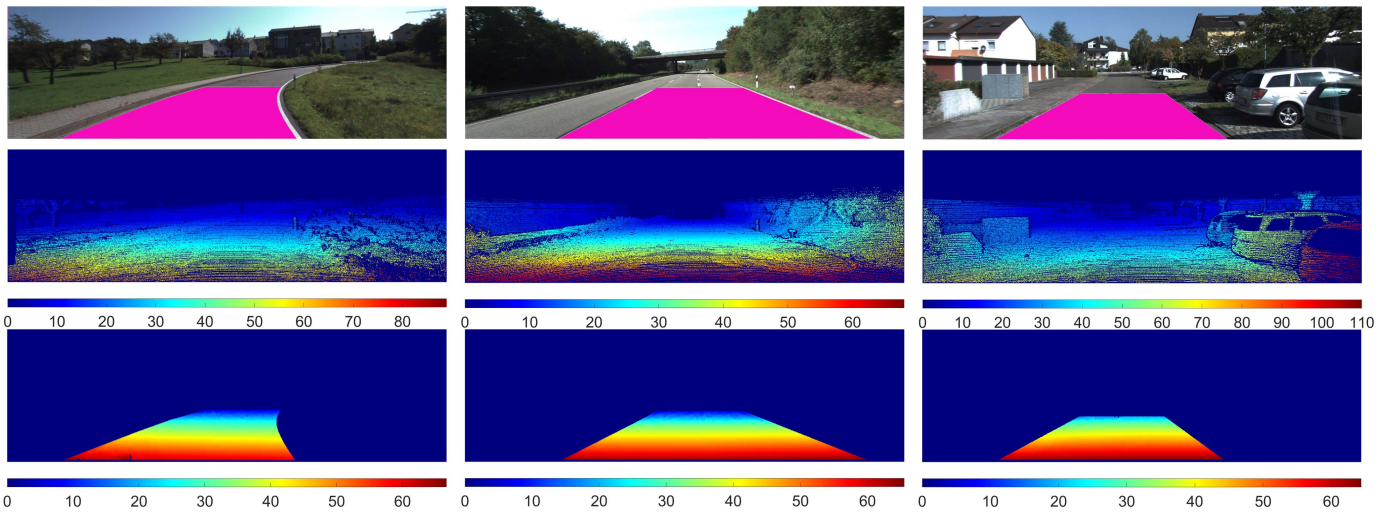


Fig. 14. Experimental results of the KITTI stereo 2012 dataset. The first row shows the left images, where areas in magenta are our manually selected road surface. The second row shows the disparity ground truth. The third row shows the results obtained from the proposed algorithm.

Furthermore, we select one row from the disparity map to evaluate the performance of subpixel enhancement and global refinement (see Fig. 13). The integer disparity d oscillates along the selected row and drops down abruptly when a discontinuity occurs. After the subpixel enhancement, the disparity d is replaced with a better one d_s between $d-1$ and $d+1$. The iterative global refinement further optimizes the subpixel disparity map. After the third iteration, the disparities change more smoothly in a continuous area but interrupt suddenly when reaching a discontinuity.

Since the datasets we create only contain the ground truth of 3D reconstruction, the KITTI stereo 2012 dataset [40] is used to further evaluate the disparity accuracy of our algorithm. Some experimental results are illustrated in Fig. 14. Due to the fact that the proposed algorithm only aims at reconstructing the road surface, we select a region of interest (see the magenta areas in the first row) from each image to evaluate the performance of our algorithm. The corresponding disparity results in the region of interest are shown in the third row. The percentage of error pixels (threshold: two pixels) is around 0.73% and the average error in pixels is about 0.51.

C. Reconstruction Evaluation

To further evaluate the accuracy of the reconstruction results, we create dataset 3 (see section VI-A for details) with three different sample models. An example of the left image is illustrated in Fig. 15 (a). The corresponding subpixel disparity map and 3D reconstruction are depicted in Fig. 15 (b) and (c), respectively. We select a rectangular region which includes one of the sample models from Fig. 15 (a), and the 3D reconstruction of this region can be seen in Fig. 15 (d). A surface $\kappa_0\ X_w + \kappa_1\ Y_w + \kappa_2\ Z_w + \kappa_3 = 0$ is fitted to four corners S_1, S_2, S_3 and S_4 of the selected region. Then, we select a set of random points $\boldsymbol{P_1}, \boldsymbol{P_2}, \ldots, \boldsymbol{P_n}$ on the surface of the model and estimate the distances between them and the fitted road surface. These random distances provide the measurement range of the model height. Similarly, the groove depth can be estimated by computing the distances between a

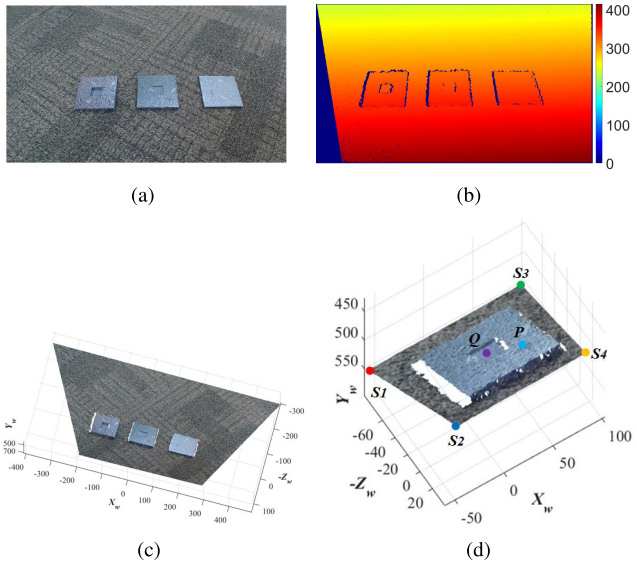


Fig. 15. Sample model 3D reconstruction. (a) left image. (b) subpixel disparity map with post-processing. (c) reconstructed scenery. (d) selected 3D point cloud which includes model B.

group of points $\boldsymbol{Q_1}, \boldsymbol{Q_2}, \ldots, \boldsymbol{Q_n}$ in a groove and the model surface. Table II details the range of the measured model height and groove depth, where D represents the approximated distance from the camera to sample models.

From Table II, the maximal absolute error of the 3D reconstruction is approximately 3 mm, and it increases slightly when D increases. The reconstruction precision is inversely proportional to the depth [41]. Furthermore, since the baseline of the ZED camera is fixed and cannot be increased to further improve the precision, we mount it to a relatively low height and it is kept as perpendicular as possible to the road surface to reduce the average depth, which guarantees a high reconstruction accuracy.

D. Processing Speed

The algorithm is implemented in C language on an Intel Core i7-4720HQ CPU (2.6 GHz) using a single thread. After

TABLE II
3D RECONSTRUCTION MEASUREMENT RANGE

Target	Measurement range (mm)				
	$D \approx 450mm$	$D \approx 470mm$	$D \approx 500mm$	$D \approx 550mm$	$D \approx 650mm$
Model A height	09.72 – 10.21	09.64 – 11.12	10.31 – 12.19	09.59 – 12.37	08.99 – 12.62
Model B height	09.86 – 10.32	09.91 – 10.47	10.07 – 11.25	10.10 – 11.99	10.86 – 12.36
Model C height	04.62 – 05.54	04.92 – 06.11	05.72 – 06.93	06.61 – 07.18	06.69 – 07.54
Groove A depth	07.77 – 08.44	08.31 – 09.54	05.92 – 09.17	05.49 – 07.26	09.37 – 11.83
Groove B depth	02.21 – 05.12	04.88 – 05.32	04.97 – 06.51	06.28 – 07.57	05.29 – 06.63

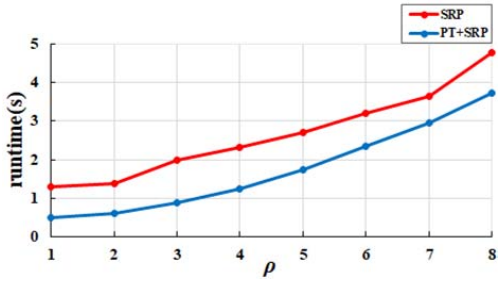


Fig. 16. Comparison between SRP and PT+SRP in terms of the runtime.

TABLE III
ALGORITHM RUNTIME

Dataset	Frames	Resolution	Runtime (s)
Dataset 1	35	1240×609	0.71
Dataset 2	35	1249×620	0.84
Dataset 3	21	2081×1048	2.23

the PT, each point on row v in the target image is shifted $a_0 + a_1 v - \delta$ pixels to obtain a reference view, which greatly reduces the search range for stereo matching. The evaluation of the PT with respect to the runtime is illustrated in Fig. 16. The PT accelerates the processing speed of the SRP stereo when using different block sizes. When $\rho = 5$, the processing speed is increased by over 36%. The runtime of different datasets is shown in Table III. Although the proposed algorithm does not run in real time, the authors believe that its speed can be increased in the future by exploiting the parallel computing architectures.

VII. CONCLUSION AND FUTURE WORK

The main novelties of this paper include PT, CMV, and disparity map global refinement. We created three datasets and made them publicly available to contribute to 3D reconstruction-based pothole detection. The PT not only enhances the similarity of a GP between two images but also reduces the search range for stereo matching. This helps the SRP stereo perform more accurately and efficiently. The CMV further offsets the insufficient propagation in the SRP stereo and guarantees the feasibility of parabola interpolation in the subpixel enhancement phase. By iteratively minimizing the energy with respect to the interpolated parabolas, the subpixel disparity map is optimized. The disparities in a continuous area become more smooth, but they are preserved when discontinuities occur. The maximal absolute error of the 3D reconstruction is around 3 mm, which satisfies the requirement of millimeter accuracy for on-road damage detection. Furthermore, due to the high precision of the proposed system, users can apply it to road surface SLAM (Simultaneous Localization and Mapping) for many smart city applications.

However, the propagation strategy in the proposed algorithm makes it difficult to fully exploit the parallel computing architecture of the graphics cards to estimate disparity maps. Therefore, we aim to come up with a more efficient SRP strategy which can be adapted for different platforms. Furthermore, errors in stereo calibration always affect the precision of the stereo matching dramatically. Hence, we aim to design a self-calibration algorithm to enhance the robustness of our proposed stereo vision system, and the reconstructed sceneries will be used for 3D pothole detection.

REFERENCES

[1] C. Koch, K. Georgieva, V. Kasireddy, B. Akinci, and P. Fieguth, "A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure," *Adv. Eng. Inform.*, vol. 29, no. 2, pp. 196–210, 2015.

[2] T. Kim and S. K. Ry, "Review and analysis of pothole detection methods," *J. Emerg. Trends Comput. Inf. Sci.*, vol. 5, no. 8, pp. 603–608, Aug. 2014.

[3] BBC News. *Councils in England Face Huge Raod Repair Bills*. Accessed: Jan. 6, 2015. [Online]. Available: http://www.bbc.co.uk/news/uk-england-30684854

[4] E. Schnebele, B. Tanyu, G. Cervone, and N. Waters, "Review of remote sensing methodologies for pavement management and assessment," *Eur. Transp. Res. Rev.*, vol. 7, no. 2, pp. 1–19, 2015.

[5] L. Cruz, L. Djalma, and V. Luiz, "Kinect and RGBD images: Challenges and applications graphics," in *Proc. 25th SIBGRAPI Conf. Patterns Images Tuts. (SIBGRAPI-T)*, 2012, pp. 36–49.

[6] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[7] F. Sadjadi and E. Ribnick, "Passive 3D sensing, and reconstruction using multi-view imaging," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2010, pp. 68–74.

[8] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.

[9] A. T. Ihler, J. W. Fischer, III, and A. S. Willsky, "Loopy belief propagation: Convergence and effects of message errors," *J. Mach. Learn. Res.*, vol. 6, pp. 905–936, May 2005.

[10] M. F. Tappen and W. T. Freeman, "Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters," in *Proc. 11th IEEE Int. Conf. Comput. Vis.*, Oct. 2003, p. 900.

[11] 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.

[12] M. G. Mozerov and J. van de Weijer, "Accurate stereo matching by two-step energy minimization," *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1153–1163, Mar. 2015.

[13] S. N. Sinha, D. Scharstein, and R. Szeliski, "Efficient high-resolution stereo matching using local plane sweeps," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1582–1589.

[14] M. Bleyer, C. Rhemann, and C. Rother, "Extracting 3D scene-consistent object proposals and depth from stereo images," in *Proc. Eur. Conf. Comput. Vis.*, Oct. 2012, pp. 467–481.

[15] K. Yamaguchi, D. McAllester, and R. Urtasun, "Efficient joint segmentation, occlusion labeling, stereo and flow estimation," in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 756–771.

[16] R. Sara, "Finding the largest unambiguous component of stereo matching," in *Proc. Eur. Conf. Comput. Vis.*, May 2002, pp. 900–914.

[17] R. Sara, "Robust correspondence recognition for computer vision," in *Proc. Comput. Statist.*, 2006, pp. 119–131.

[18] J. Cech and R. Sara, "Efficient sampling of disparity space for fast and accurate matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2007, pp. 1–8.

[19] R. Spangenberg, T. Langner, and R. Rojas, "Weighted semi-global matching and center-symmetric census transform for robust driver assistance," in *Proc. Int. Conf. Comput. Anal. Images Patterns*, 2013, pp. 34–41.

[20] O. Miksik, Y. Amar, V. Vineet, P. Pérez, and P. H. Torr, "Incremental dense multi-modal 3d scene reconstruction," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Jun. 2015, pp. 908–915.

[21] S. Pillai, S. Ramalingam, and J. J. Leonard, "High-performance and tunable stereo reconstruction," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Nov. 2016, pp. 3188–3195.

[22] Z. Zhang, X. Ai, and N. Dahnoun, "Efficient disparity calculation based on stereo vision with ground obstacle assumption," in *Proc. 21st Eur. Signal Process. Conf. (EUSIPCO)*, 2013, pp. 1–5.

[23] R. Fan and N. Dahnoun, "Real-time implementation of stereo vision based on optimized normalized cross-correlation and propagated search range on a GPU," in *Proc. IEEE Int. Conf. Imag. Syst. Technol. (IST)*, Nov. 2017, pp. 1–6.

[24] Q. Yang, L. Wang, R. Yang, H. Stewénius, and D. Nistér, "Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 3, pp. 492–504, Mar. 2009.

[25] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 504–511, Feb. 2013.

[26] Z. Zhang, X. Ai, N. Canagarajah, and N. Dahnoun, "Local stereo disparity estimation with novel cost aggregation for sub-pixel accuracy improvement in automotive applications," in *Proc. IEEE Intell. Veh. Symp. (IV)*, Jun. 2012, pp. 99–104.

[27] H. Hattori and A. Maki, "Stereo without depth search and metric calibration," in *Proc. IEEE Conf Comput. Vis. Pattern Recognit.*, vol. 1. Jun. 2000, pp. 177–184.

[28] H. Nakai, N. Takeda, H. Hattori, Y. Okamoto, and K. Onoguchi, "A practical stereo scheme for obstacle detection in automotive use," in *Proc. 17th Int. Conf. Pattern Recognit. (ICPR)*, vol. 3. 2004, pp. 346–350.

[29] Z. Hu, F. Lamosa, and K. Uchimura, "A complete uv-disparity study for stereovision based 3d driving environment analysis," in *Proc. 15th Int. Conf. 3-D Digit. Imag. Modeling (DIM)*, 2005, pp. 204–211.

[30] S. Leutenegger, M. Chli, and R. Y. Siegwart, "BRISK: Binary robust invariant scalable keypoints," in *Proc. Int. Conf. Comput. Vis.*, 2011, pp. 2548–2555.

[31] S. Roy, "Stereo without epipolar lines: A maximum-flow formulation," *Int. J. Comput. Vis.*, vol. 34, nos. 2–3, pp. 147–161, 1999.

[32] I. Haller and S. Nedevschi, "Design of interpolation functions for subpixel-accuracy stereo-vision systems," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 889–898, Feb. 2012.

[33] A. Blake, P. Kohli, and C. Rother, *Markov Random Fields for Vision and Image Processing*. Cambridge, MA, USA: MIT Press, 2011.

[34] S. Z. Li, *Center for Biometrics and Security Research & National Laboratory of Pattern Recognition*. Beijing, China: Institute of Automation Chinese Academy of Science, 2012.

[35] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. 6th Int. Conf. Comput. Vis.*, 1998, pp. 839–846.

[36] R. Szeliski *et al.*, "A comparative study of energy minimization methods for Markov random fields with smoothness-based priors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 6, pp. 1068–1080, Jun. 2008.

[37] U. Ozgunalp, R. Fan, X. Ai, and N. Dahnoun, "Multiple lane detection algorithm based on novel dense vanishing point estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 3, pp. 621–632, Mar. 2017.

[38] G. G. Slabaugh, "Computing euler angles from a rotation matrix," *Retrieved August*, vol. 6, no. 2000, pp. 39–63, 1999.

[39] STEREOLABS. *Stereolabs Products*. Accessed: May 29, 2017. [Online]. Available: https://www.stereolabs.com/zed/specs/

[40] A. Andreas, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3354–3361.

[41] D. F. Llorca, M. A. Sotelo, I. Parra, M. Ocaña, and L. M. Bergasa, "Error analysis in a stereo vision-based pedestrian detection sensor for collision avoidance applications," *Sensors*, vol. 10, no. 4, pp. 3741–3758, 2010.

Rui Fan (GS'16) received the B.Sc. degree in control science and engineering from the Harbin Institute of Technology in 2015. He is currently pursuing the Ph.D. degree with the Visual Information Laboratory, University of Bristol. His research interests include multi-view geometry, real-time depth measurement, high-performance computing, and automotive applications, e.g., lane detection, obstacle detection, pothole detection, and visual tracking.

Xiao Ai received the B.Sc. and Ph.D. degrees in electrical and electronics engineering from the University of Bristol, Bristol, U.K., in 2007 and 2012, respectively. He is a Post-Doctoral Researcher with the University of Bristol. His Ph.D. was specialized in 3D imaging techniques and applications. His current research interests include embedded real-time signal processing, optoelectronics for spaceborne remote sensing, and automotive obstacle detection applications. He also has extensive experience in machine vision.

Naim Dahnoun received the Ph.D. degree in biomedical engineering from the University of Leicester, Leicester, U.K., in 1990. He was with the Leicester Royal Infirmary as a Researcher on blood flow measurements for femoral bypass grafts and then with the University of Leicester as a Lecturer in digital signal processing (DSP). In 1993, he started a new research in optical communication at The University of Manchester Institute of Science and Technology, Manchester, U.K., on wideband optical communication links before joining the Department of Electrical and Electronic Engineering, University of Bristol, Bristol, U.K., where he is a Reader in learning and teaching of DSP in 1994. His main research interests include real-time digital signal processing applied to biomedical engineering, video surveillance, automotive, and optics. In 2003, in recognition of the important role played by universities in educating engineers in new technologies such as real-time DSP, he received the first Texas Instruments DSP Educator Award from Texas Instruments (NYSE:TXN) for his outstanding contributions to furthering education in DSP technology.