IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 25, 2019, accepted April 1, 2019, date of publication April 11, 2019, date of current version April 22, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2910194

An Efficient Dense Stereo Matching Method for Planetary Rover

HAICHAO LI, LIANG CHEN, AND FENG LI, (Senior Member, IEEE)
Qian Xuesen Laboratory of Space Technology, China Academy of Space Technology, Beijing 100094, China

Corresponding author: Haichao Li (lihaichao2000@163.com)

The work of H. Li was supported by the National Nature Science Foundation of China under Grant 61773383. The work of L. Chen was supported by the National Nature Science Foundation of China under Grant 41601483.

ABSTRACT Stereo matching is one of the most important and challenging subjects in the field of planetary rover with a stereo vision system. The generated disparity map can make rover to avoid the obstacle and explore the planetary surface automatically. In this paper, we propose an efficient dense stereo matching method to generate disparity maps for a planetary rover, which relies on 3-D plane fitting, adaptive penalties, and coarse-to-fine disparity constraint. In order to achieve efficient stereo matching at the coarsest level of the pyramid, we present a 3-D plane fitting to reduce the disparity search range and propose adaptive penalties in the more-global matching method for obtaining an accurate disparity map. At the finer level, our method then estimates the disparity search range based on coarse-to-fine disparity constraint and utilizes adaptive penalties to obtain an accurate disparity map. The extensive experiments with stereo images of Chang'e-3 rover demonstrate that our approach can generate disparity maps efficiently and accurately compared with the most state-of-the-art semi-global matching methods, especially at low texture regions, depth discontinuities, and occlusion regions.

INDEX TERMS Stereo matching, disparity map, coarse-to-fine architecture, 3-D plane, planetary rover.

I. INTRODUCTION

Planetary exploration is always an ambitious mission which launches a series of spacecraft to land on the planet surface. It will carry rovers to explore the planetary surface and carry out science missions. Therefore, the planetary rover is an important component, and it needs to complete many higher level tasks in an unstructured environment, such as obstacle detection, environmental perception, path planning, navigation, localization and so on. Among the various tasks, the ability to recognize the reliable three-dimensional (3D) terrain information surrounding the rover is essential, which can guide the rover to perform obstacle avoidance and path planning to reach the required destinations [1]. Furthermore, because of the significant delay in communication between the Earth and the planets, future planetary rovers must have the ability to perceive the environment independently in order to make them run in unstructured environment safely and autonomously [2].

The stereo vision and the scanning LiDAR are two kinds of approaches to achieve reliable 3D terrain information. Based on the stereo vision technology, a 3D map can be generated from stereo images taken by cameras mounted on the planetary rover. Stereo vision is an attractive technology for rover navigation because it is passive and only a small amount of power is required for imaging electronics [3]. Scanning LiDAR technology can offer much longer operational range (usually hundreds of meters) and higher accuracy than stereo vision technology (usually tens of meters). However, this comes with an additional mass and power requirements, and also increases launch costs.

A successfully application of 3D navigation in unstructured environment is the Sojourner rover of Mars Pathfinder [4], [5] with two stereo cameras and five laser stripe emitters to detect obstacles. The Mars Exploration Rovers (MER) Spirit and Opportunity have the ability to navigate independently through unknown environment, which also use passive stereo vision to detect potential terrain hazards [6]. For the Mars Science Laboratory Curiosity rover, stereo images of the Martian surface can be acquired by the Curiosity rover Mast Camera system [7], and 3D information can be achieved for obstacle avoidance, plan planning, and navigation and positioning.

A. RELATED WORK

In general, the stereo matching methods can be categorized into local, global, and semi-global matching methods.

The associate editor coordinating the review of this manuscript and approving it for publication was You Yang.

2169-3536 © 2019 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Firstly, local matching methods have low time complexities which are widely used for planetary rover with limited computational resources and power. Local methods mainly include feature-based and area-based matching methods. The feature-based methods get only sparse points, and are hence insufficient for reconstructing detailed 3D terrain. While the area-based methods can generate a dense disparity map, which is more suitable for 3D perception of a planetary rover. The most common methods used mainly include sum of absolute differences (SAD) [8], [9], normalized cross correlation (NCC) [10], and correlation coefficient [6]. Although local methods have been widely utilized in planetary rovers, there may be a potentially large amount of erroneous matches due to lack of texture, occlusions, and repetitive patterns on the planetary surface [11]. To address the matching ambiguity caused by local methods, several multi-resolution stereo approaches have been proposed which use the disparity computed at a coarser scale to restrict the search range at the current scale [12], [13].

Secondly, compared to local methods, global matching methods based on energy minimization can address matching ambiguities more effectively and can achieve more precise results especially in disparity discontinuous and textureless regions. Global methods [14] usually formulate the stereo matching problem as the minimization over an energy function defined on the Markov Random Field (MRF) model where a smoothness assumption is made to achieve better results but with a higher computational cost [15]. There are many different strategies to find the minimum of the global energy function, such as graph cuts (GC) [16], dynamic programming (DP) [17], and belief propagation (BF) [18]. Due to the high computational cost, few researches have been conducted on global matching for planetary rovers. A global matching method has been published based on feature points and MRF model [19] for planetary images. To reduce the computational cost, a hierarchical structure approach is proposed to reduce the computational complexity of the GC algorithm [20].

Thirdly, local methods usually fails to match at ambiguous low texture areas, while global methods require more processing time although they have high matching accuracy. The Semi-Global Matching (SGM) method [21], [22], incorporating the advantages of both local and global methods, provides a good trade-off between accuracy and complexity. The Planetary Robotics 3D viewer (Pro3D) is developed to visualize stereo images collected by rovers on the Martian surface [23], and the 3D models are reconstructed using the SGM method based on the acquired numerous stereo images using both MER scientific Pancam system and navigational NavCam system [24]. For reducing computational complexity, two SGM methods based on the coarse-to-fine strategy are proposed in [25], [26]. Moreover, original or variant methods of SGM have been implemented on different computing architectures such as Graphics Processing Unit (GPU) [27], Field-Programmable Gate Array (FPGA) [28], and other embedded devices [29]. Although the SGM method has been widely used, this method may yield the known streaking artifacts. The More-Global Matching (MGM) method aims at improving the accuracy of SGM method according to a more global strategy, which removes the streaking artifacts of SGM, and improves the quality of the disparity result [30]. However, all the SGM and MGM methods are performed with a full disparity search range, which requires high memory.

B. CONTRIBUTIONS OF OUR WORK

For the left and right images obtained by a planetary rover, the disparity search range is unknown in advance, but it is a very important parameter due to the limited computing resources, storage resources, and energy resources of the rover. Under the unknown unstructured environment of planetary surface, it is difficult for us to get and set a fixed search range just like in the structured environment. Therefore, if we set a large disparity search range, the oversize of the search range will consume more time, take up more memory space, and suffer from matching ambiguities especially for low texture areas. If the search range is set too small, the correct disparity is often not available. Therefore, a reliable estimation of disparity search range is the prerequisite of fast stereo matching for a planetary rover, as well as the requirement for saving computation time and memory space. Furthermore, for rover images of planetary surfaces, in addition to low texture at a planetary surface, disparity discontinuities usually exist along the boundaries of prominent features such as large rocks.

In this paper, our point of interest is to generate disparity map efficiently and accurately with stereo images of planetary surface, in order to meet the requirements of timeliness and reliability of the planetary rover. The major contributions of this work are as follows:

(1) The proposed method builds a coarse-to-fine pyramid framework, through which the pixel-wise disparity search range at a finer level is adaptively restricted to a few pixels based on the disparity map at the adjacent coarser level.

(2) According to the imaging characteristics of planetary rover from near to far on a planetary surface, the disparity value is approximately linear changing with the distance between planetary rover and planetary surface, which can be validated by reconstructed disparity maps (see Fig. 1). Therefore, we propose a method of 3-D plane fitting to automatically determine pixel-wise disparity search range at the coarsest level to reduce the computational complexity.

(3) To improve the disparity accuracy of state-of-the-art semi-global matching methods especially at disparity discontinuities and occlusion regions, we propose a strategy to adaptively determine the penalties in cost aggregation according to the extracted edges.

The remainder of the paper is organized as follows. The details of the proposed method are presented in section II. Section III gives the experimental results and analysis. Section IV is the conclusion.

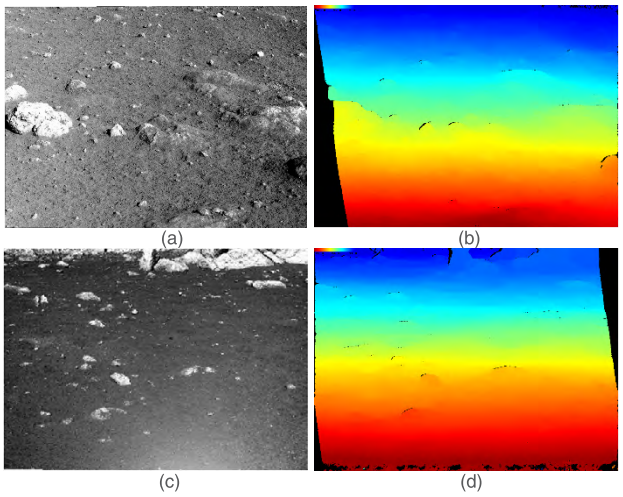

FIGURE 1. Disparity maps of lunar stereo images taken by Chang'e-3 Yutu rover. (a) Left image 1. (b) Disparity map of (a), with disparity range from −106 to 2 pixels. (c) Left image 2. (d) Disparity map of (c), with disparity range from −23 to 71 pixels.

II. PROPOSED STEREO MATCHNG

The input of our proposed method is two rectified left and right images obtained with epipolar rectification method [31]. A stereo pair of Gaussian pyramids [32] with L levels are constructed from the two rectified images. At level l=1, respective images are the input rectified images, and at each level $l \in L, \cdots, 2$, respective images are obtained by halving the resolution of the corresponding images at level $l-1$. The framework overview of the proposed method is shown in Fig. 2.

At the coarsest level $l = L$, the disparity map is generated by a 3-step scheme, which consists of (1) 3-D plane fitting, (2) cost computation and aggregation with adaptive penalties, and (3) post processing. The complete processing scheme at the coarsest level provides a robust and effective disparity estimation based on 3-D plane fitting, which can greatly reduce the computational complexity by restricting the disparity search range.

Further, from level $l = L-1$ to level $l = 1$, the disparity map generated at coarser level will guide disparity generation at finer level. When the process comes to the finest level ($l = 1$), we will obtain disparity maps of the input rectified images. At a finer level, however, a scheme is proposed which using coarse-to-fine disparity constraint instead of 3-D plane fitting to reduce the disparity search range. That is, the disparity map can be generated by a 3-step scheme which consists of (1) coarse-to-fine disparity constraint, (2) cost computation and aggregation with adaptive penalties, and (3) post processing.

In the following sections, we will discuss several key components in detail.

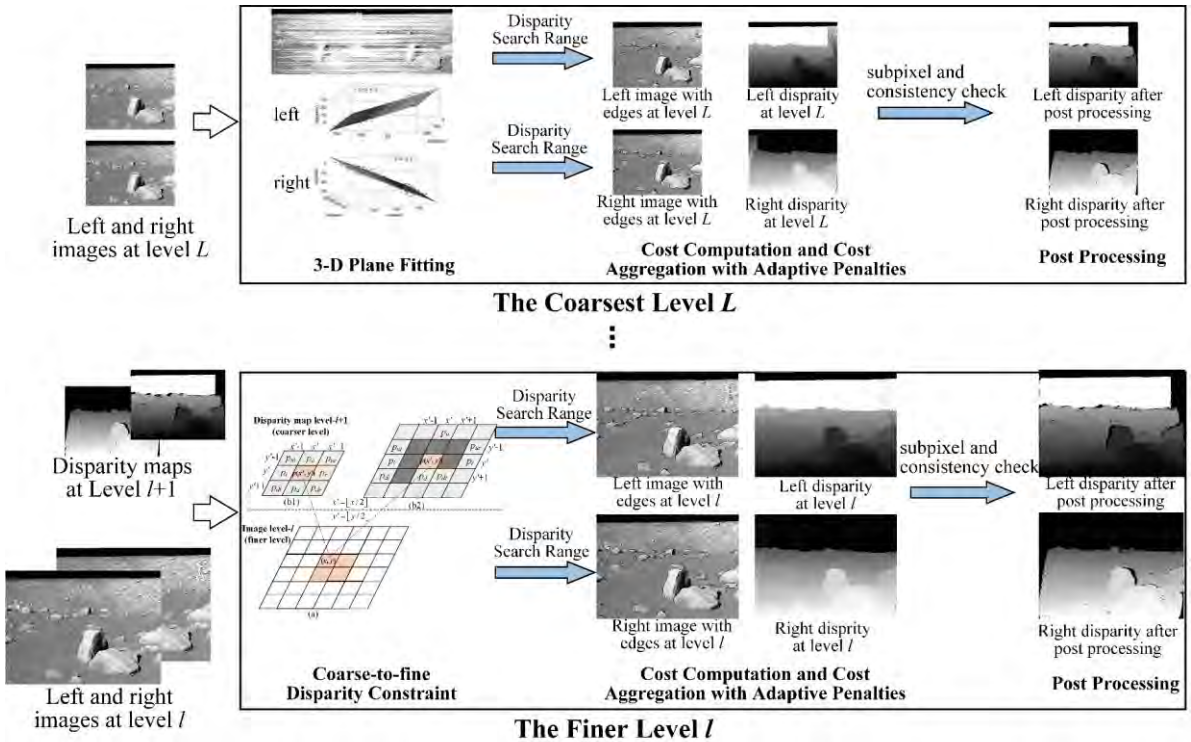

FIGURE 2. The framework overview of the proposed method. The upper row shows the complete disparity map generation steps at the coarsest level *L*, including 1) 3-D plane fitting. 2) Cost computation and aggregation with adaptive penalties. 3) Post processing, while the lower row shows the disparity map generation steps at the other levels, including 1) coarse-to-fine disparity constraint, 2) cost computation and aggregation with adaptive penalties, and 3) post processing. Note that 3-D plane fitting is only applied at the coarsest level, however, the coarse-to-fine disparity constraint is instead of 3-D plane fitting at finer levels.

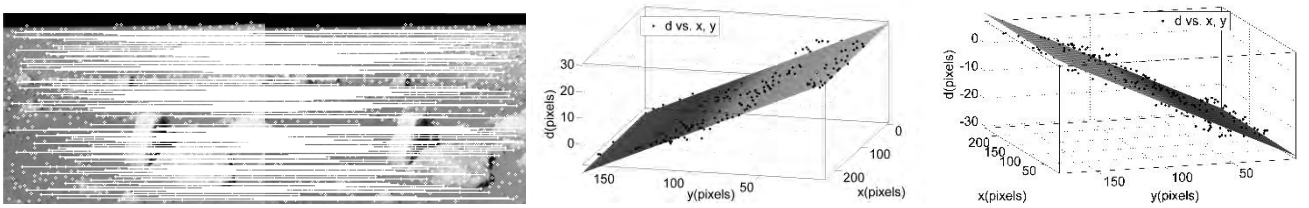

FIGURE 3. The matched 260 pairs of SIFT keypoints (left). The fitted 3-D plane for the left image (middle), a = −0.00517, b = −0.227, c = 30.5. The fitted 3-D plane for the right image (right), a = 0.0034, b = −0.2268, and c = −30.3539.

A. 3-D PLANE FITTING AT THE COARSEST LEVEL

Scale invariant feature transform (SIFT) features [33] are used to detect keypoints from the left and right images at the coarsest level. The detected keypoints are matched through the nearest neighbor approach as proposed in [33], where the epipolar constraint and random sample consensus (RANSAC) algorithm [34] are also employed to effectively eliminate the mismatched pairs. Suppose two sets of n matched keypoints $\{(x_i, y_i)\}$ and $\{(x'_i, y'_i)\}$ have been obtained, which belonging to the left and right images, respectively, $i = 1, 2, \cdots, n, n \geq 3$. The pixel-wise disparity search range for the two coarsest images is determined based on the 3-D plane fitting using the following steps:

Step 1: Construct two sets of 3-D point coordinates $\{(x_i, y_i, d_i)\}$ and $\{(x'_i, y'_i, d'_i)\}$, where $d_i = x'_i - x_i$ and $d'_i = x_i - x'_i$.

Step 2: For the set $\{(x_i, y_i, d_i)\}$ obtained from the left image, a 3-D plane, $d = ax + by + c$, can be fitted with the least square method as shown in (1).

$$S = \sum_{i=1}^{n} (ax_i + by_i + c - d_i)^2 \tag{1}$$

where a, b, and c are the 3-D plane coefficients to be estimated.

Step 3: For the keypoint set $\{(x_i, y_i)\}$, compute the disparity set $\{D_i\}$, $D_i = ax_i + by_i + c$. Then compute the disparity difference set $\{\Delta d_i\}$, $\Delta d_i = D_i - d_i$, in which the minimum and maximum values are expressed as $D_{\max}$ and $D_{\min}$, respectively.

Step 4: For every pixel (x, y) in the left image at the coarsest level, calculate the disparity value with the 3-D plane by (2).

$$d(x, y) = ax + by + c \tag{2}$$

Here, the minimum disparity $d_{\min}(x, y)$ and the maximum disparity $d_{\max}(x, y)$ of pixel (x, y) are computed by (3).

$$\begin{cases} d_{\min}(x, y) = d(x, y) + D_{\min} - \Delta d \\ d_{\max}(x, y) = d(x, y) + D_{\max} + \Delta d \end{cases} \tag{3}$$

where Δd is a given disparity search offset at the coarsest level. Therefore, the disparity search range of pixel (x, y) can be expressed as $[d_{\min}(x, y), d_{\max}(x, y)]$.

Step 5: Similarly, repeating the above steps 1∼ 4 for another set $\{(x'_i, y'_i, d'_i)\}$, we can determine pixel-wise disparity search range in the right image, $[d'_{\min}(x, y), d'_{\max}(x, y)]$.

Fig. 3 gives an example of the matched SIFT keypoints and the fitted 3-D plane for the left and right images.

B. COST COMPUTATION AND AGGREGATION WITH ADAPTIVE PENALTIES

MGM and SGM for dense stereo matching are efficient methods to perform approximate energy minimization for 2D MRF, which are the state-of-the-art semi-global matching methods. However, these methods attempt to find the best matching point by traversing a full disparity search range. Again, there possibly be inaccurate disparity generated when utilizing the constant penalties especially at depth discontinuities and occluded regions. Several semi-global matching methods have been proposed to adjust the penalties based on the intensity gradient between neighboring two pixels [22], [35]. However, the pixel with intensity changes may not correspond to disparity discontinuity, and at the same time, the disparity is easy to be seriously interfered by image noise.

In general, the disparity of the same object surface changes smoothly, while the disparity at discontinuous boundaries are mostly distributed on the edges of object. We apply the Canny detector [36] to detect edges from the left and right images at level l, $l \in L, \cdots, 1$, and then the penalties are adaptively determined based on the extracted edges to improve the disparity accuracy especially at disparity discontinuities, obstacle edge areas, and occlusion regions.

1) COST COMPUTATION AND AGGREGATION

For the computation of matching cost, the census transform [37] is often a preferred choice, because it has the advantage of being independent of luminosity and contrast differences between cameras. For a pixel $\mathbf{p}(x, y)$ in the left image at level l, the matching cost $C(\mathbf{p}, d)$ at disparity d is computed based on census transform, $d \in [d_{\min}(x, y), d_{\max}(x, y)]$.

However, the pixel-wise matching only with $C(\mathbf{p}, d)$ is generally ambiguous. Therefore, a smoothness term is added by penalizing the changes of neighboring disparities. Considering the NP-complete problem in 2D, matching costs are aggregated in 1D from all directions. For each propagation

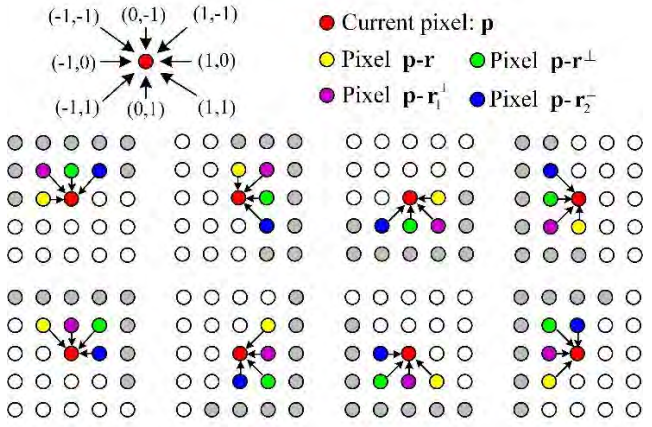

FIGURE 4. The eight path directions r and the corresponding recursion directions R for each r.

direction $\mathbf{r}$, the path cost $L_{\mathbf{r}}(\mathbf{p}, d)$ of the pixel $\mathbf{p}$ at disparity d is computed recursively by (4).

$$\begin{aligned} &L_{\mathbf{r}}(\mathbf{p}, d) \\ &\quad = C(\mathbf{p}, d) \\ &\qquad + \frac{1}{4} \sum_{\mathbf{R}\in\{\mathbf{r},\mathbf{r}^{\perp},\mathbf{r}_1^{\perp},\mathbf{r}_2^{\perp}\}} \min \begin{pmatrix} L_{\mathbf{r}}(\mathbf{p}-\mathbf{R}, d), \\ L_{\mathbf{r}}(\mathbf{p}-\mathbf{R}, d+1) + P_{1,\mathbf{R}}, \\ L_{\mathbf{r}}(\mathbf{p}-\mathbf{R}, d-1) + P_{1,\mathbf{R}}, \\ \min\limits_{d_i} L_{\mathbf{r}}(\mathbf{p}-\mathbf{R}, d_i) + P_{2,\mathbf{R}} \end{pmatrix} \\ &\qquad - \min_{d_i} L_{\mathbf{r}}(\mathbf{p}-\mathbf{R}, d_i) \end{aligned} \tag{4}$$

Here d_i belongs to disparity search range of pixel $\mathbf{p}-\mathbf{R}$, which is computed by 3-D plane fitting when $l = L$, otherwise is computed from disparity at a coarser level $l+1$ when $l \in L1, \cdots, 1$. $P_{1,\mathbf{R}}$ and $P_{2,\mathbf{R}}$(with $0 < P_{1,\mathbf{R}} < P_{2,\mathbf{R}}$) are two penalties which can be adaptively determined (see the following Section) for penalizing disparity discontinuities of neighboring pixels. $\mathbf{R}$ represents recursion directions from path direction $\mathbf{r}$, which is expressed as $\mathbf{R} \in \{\mathbf{r}, \mathbf{r}^{\perp}, \mathbf{r}_1^{\perp}, \mathbf{r}_2^{\perp}\}$. As shown in Fig. 4, this paper considers eight path directions, $\mathbf{r}\in\{(-1, 0), (1, 0), (0, 1), (0, -1), (-1, -1), (1, -1), (1, 1), (-1, 1)\}$, therefore, the corresponding direction respectively expressed as $\mathbf{R} \in \{(-1, 0), (0, -1), (-1, -1), (1, -1)\}$, $\mathbf{R} \in \{(1, 0), (0, 1), (1, 1), (-1, 1)\}$, $\mathbf{R} \in \{(0, 1), (-1, 0), (-1, 1), (-1, -1)\}$, $\mathbf{R} \in \{(0, -1), (1, 0), (1, -1), (1, 1)\}$, $\mathbf{R} \in \{(-1, -1), (1, -1), (0, -1), (1, 0)\}$, $\mathbf{R} \in \{(1, -1), (1, 1), (1, 0), (0, 1)\}$, $\mathbf{R} \in \{(1, 1), (-1, 1), (0, 1), (-1, 0)\}$, $\mathbf{R} \in \{(-1, 1), (-1, -1), (-1, 0), (0, -1)\}$. The minimum path cost $\min_{d_i} L_{\mathbf{r}}(\mathrm{p} - \mathbf{R}, d_i)$ of the previous pixel is subtracted from the whole term.

The aggregated cost $S(\mathbf{p}, d)$ for each pixel $\mathbf{p}$ with disparity d is calculated by summing the path costs in all directions $\mathbf{r}$ as presented in (5). The disparity $d(\mathbf{p})$ for a pixel $\mathbf{p}$ is then calculated by applying Winner-Take-All (WTA) strategy to $S(\mathbf{p}, d)$ as shown in (6).

$$S(\mathbf{p}, d) = \sum_{\mathbf{r}} L_{\mathbf{r}}(\mathbf{p}, d) \tag{5}$$

$$d(\mathbf{p}) = \arg \min_{d\in[d_{\min}(x,y), d_{\max}(x,y)]} S(\mathbf{p}, d) \tag{6}$$

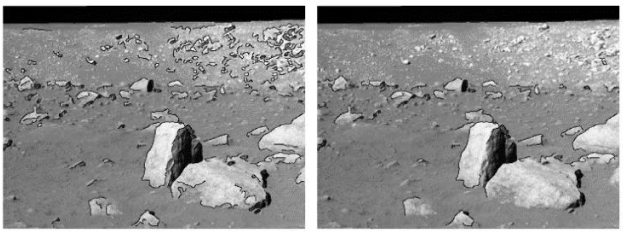

FIGURE 5. Example of Canny edges in left image at level $l = 2$ (with 3 levels pyramid). (a) The detected 167 edges. (b) The remaining 50 edges after processing.

2) ADAPTIVE PENALTIES BASED ON EDGES

Our proposed method uses adaptive penalties based on detected Canny edges instead of constant penalties. Suppose the detected edges in the left image at level l are expressed as $\{e_1, \cdots, e_N\}$, and the lengths of the corresponding edges are $\{s_1, \cdots, s_N\}$, where N is the number of the edges. The edge points on edge e_i are expressed as $e_i = \{(x_j, y_j)\}_{j=1}^{s_i}$, $i = 1, \cdots, N$. Fig. 5(a) is the edges detected with Canny detector, using a 5×5 Gaussian filter with $\sigma = 1.6$, the high threshold value 0.85 and the low threshold value 0.6. It can be seen that correct edges are detected, however, there are also edges in the flat or near flat regions with rich textures which are not really corresponding to the obvious disparity discontinuities. In our method, to eliminate the influence of edges in the areas where the disparity is not obvious changed, we further remove some edges which are not corresponding to depth discontinuities based on the obtained disparity search range of edge points, and the process is as follows:

(1) For every edge point (x_j, y_j) on edge e_i, compute the disparity range by (7).

$$\Delta d_j = d_{\max}(x_j, y_j) - d_{\min}(x_j, y_j) \tag{7}$$

where $[d_{\min}(x_j, y_j), d_{\max}(x_j, y_j)]$ is the disparity search range of point (x_j, y_j).

(2) Compute the average disparity range Δd_{e_i} of edge e_i based on all s_i edge points as shown in (8).

$$\Delta d_{e_i} = \sum\nolimits_{j=1}^{s_i} \Delta d_j \Big/ s_i \tag{8}$$

(3) If $\Delta d_{e_i} < \Delta T$, we remove edge e_i, where ΔT is a set disparity threshold, and it is set to $\Delta T = 1.5\times(L+1-l)$ according to experience.

(4) After all N edges are processed, we then obtain the remaining $\hat{N}$ edges, expressed as $\{\hat{e}_1, \cdots, \hat{e}_{\hat{N}}\}$.

Fig. 5(b) shows the remaining edges, where the edges in the flat or near flat regions are almost removed.

Then, the penalties are adaptively determined based on the remaining edges. If $\mathbf{p}$ and $\mathbf{p} - \mathbf{R}$ are both edge points, i.e., $\mathbf{p}-\mathbf{R} \in \{\hat{e}_i\}_{i=1}^{\hat{N}}$ and$\mathbf{p} \in \{\hat{e}_i\}_{i=1}^{\hat{N}}$, or both not edge points, i.e., $\mathbf{p} - \mathbf{R} \notin \{\hat{e}_i\}_{i=1}^{\hat{N}}$ and $\mathbf{p} \notin \{\hat{e}_i\}_{i=1}^{\hat{N}}$, the two penalties of image pixel $\mathbf{p}$ in (4), $P_{1,\mathbf{R}}$ and $P_{2,\mathbf{R}}$, will take the same values as the penalties P_1 and P_2 in MGM [30]. Otherwise, the two

penalties will be one tenth of P_1 and P_2 as shown in (9).

$$\begin{cases} P_{1,\mathbf{R}} = P_1/10 \\ P_{2,\mathbf{R}} = P_2/10 \end{cases} \tag{9}$$

C. COARSE-TO-FINE DISPARITY CONSTRAINT

From level $l = L - 1$ to level $l = 1$, we take a coarse-to-fine disparity constraint strategy to generate disparity map efficiently. In the coarse-to-fine scheme, the generated disparity map at a coarser level is used as prior to restrict disparity search range at a finer level.

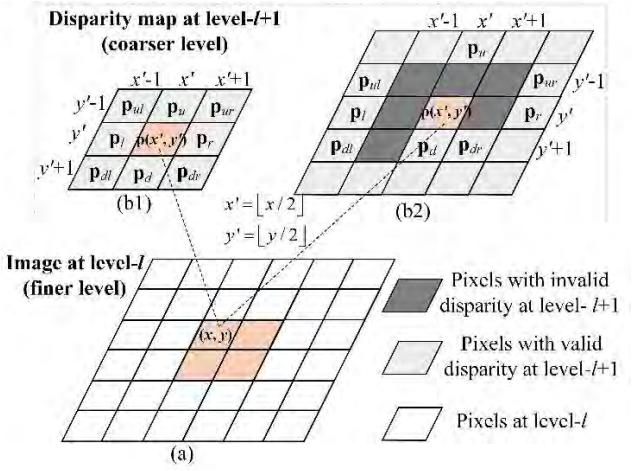

FIGURE 6. The coarse-to-fine disparity constraint between two consecutive levels. (a) A finer level l for disparity search range estimation. Relative pixels at coarser level $l + 1$ (b1) without invalid disparities, (b2) with invalid disparities.

Fig. 6 shows the coarse-to-fine disparity constraint strategy, through which the pixel-wise disparity search range at a finer level (level-l) can be determined by the disparity map at a coarser level (level-$l+1$), $l = L-1, \cdots, 1$, and the detailed process is summarized in Algorithm 1.

The proposed coarse-to-fine disparity constraint strategy has several advantages. Firstly, it has significant efficiency improvement over the traditional semi-global matching methods because pixel-wise disparity search range is only restricted to several pixels. Even though for the pixel with invalid disparity at coarser level-$l + 1$, the proposed method makes full use of adjacent pixels with valid disparities, which also restrict disparity search range at level-l to a few pixels. Furthermore, the matching ambiguities caused by large search range in complex scenes, such as textureless regions, repetitive patterns, and occlusions, can be resolved when the search range is restricted to a very short range.

D. POST PROCESSING

As shown in Fig. 2, at each level l, $l \in L, \cdots, 1$, the left disparity map can be computed by considering the left image as a reference image, while the right disparity map can be computed in a similar symmetrical manner. Then, these two disparity maps are refined by post processing. Firstly,

Algorithm 1 Adaptively Determine Pixel-Wise Disparity Search Range in the Left Image at Level l, $l \in L-1, \cdots, 1$

Input: The left image at level-l with $X \times Y$ pixels. Disparity map of the left image at level-$l+1$.

Output: The determined pixel-wise disparity search range in the left image at finer level-l.

1. **for** $x = 0$ to $X - 1$ do
2. **for** $y = 0$ to $Y - 1$ do
   2.1 For pixel coordinates (x, y) in the left image at level-l, calculate the corresponding pixel coordinates (x', y') at coarser level-$l+1$, $x' = \lfloor x/2 \rfloor$ and $y' = \lfloor y/2 \rfloor$.
   2.2 **for each** $j \in [-1, 0, 1]$ do
   From row-$y' + j$ at level-$l + 1$, we find the left and right pixels with valid disparities.
   end for j
   In the rows-$y' - 1$, y', and $y' + 1$, the left and right pixels with valid disparities are respectively expressed as $\mathbf{p}_{ul}$ and $\mathbf{p}_{ur}$, $\mathbf{p}_l$ and $\mathbf{p}_r$, and $\mathbf{p}_{dl}$ and $\mathbf{p}_{dr}$ [shown in Fig. 6(b1) or (b2)].
   2.3 Based on the pixel $\mathbf{p}(x', y')$ along column-x' at level-$l + 1$, we find the up and down pixels having valid disparities, expressed as $\mathbf{p}_u$ and $\mathbf{p}_d$ [shown in Fig. 6(b1) or (b2)].
   2.4 Among all the disparity values corresponding to $\mathbf{p}_{ul}$, $\mathbf{p}_{ur}$, $\mathbf{p}_l$, $\mathbf{p}_r$, $\mathbf{p}_{dl}$, $\mathbf{p}_{dr}$, $\mathbf{p}_u$, and $\mathbf{p}_d$, the minimum value and maximum value are expressed as $d'_{\min}$ and $d'_{\max}$, respectively. Then the disparity search range $[d_{\min}(x, y), d_{\max}(x, y)]$ of pixel (x, y) at level-l can be defined as (10).

$$\begin{cases} d_{\min}(x, y) = 2d'_{\min} - \Delta D \\ d_{\max}(x, y) = 2d'_{\max} + \Delta D \end{cases} \tag{10}$$

   where ΔD is a given disparity search offset.
3. **end for** y
4. **end for** x

we refine our disparity using subpixel interpolation [38], which is the same as that of MGM algorithm [30]. Secondly, let $\mathrm{D}^l_{\text{left}}$ and $\mathrm{D}^l_{\text{right}}$ denote the left and right subpixel disparity maps at level l, respectively. For the left-right consistency check [35], let $\mathbf{p}_\mathrm{d}$ denote the corresponding pixel in the right image which matches pixel $\mathbf{p}$ in the left image. If the two pixels in the left and right subpixel disparities satisfy (11), these pixels are marked as valid pixels.

$$\left|\mathrm{D}^l_{\text{left}}(\mathbf{p})\text{-}\mathrm{D}^l_{\text{right}}(\mathbf{p}_\mathrm{d})\right| < 1 \tag{11}$$

Otherwise, these pixels are marked as invalid pixels, which should be removed as outliers.

Fig. 7 shows the disparity maps at the coarsest level $l = 3$ (Fig. 7(a)) and level $l = 2$ (Fig. 7(b)) after post processing. For further showing the disparity correctness in depth discontinuities and occlusion regions, the disparity maps added with Canny edges are shown in Fig. 7 (right column). It is

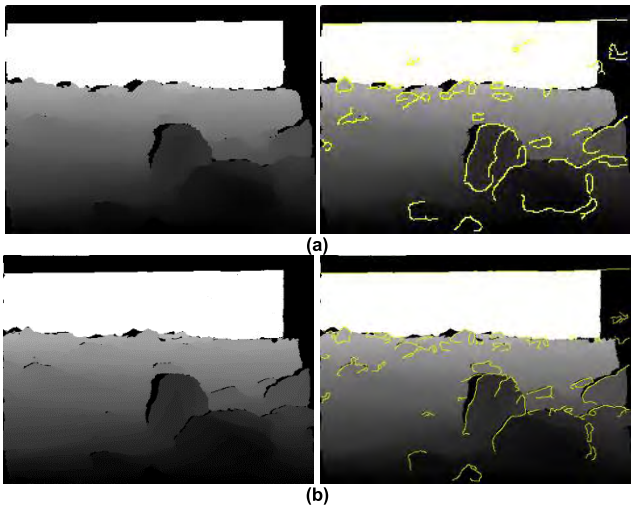

FIGURE 7. Examples of disparity map (left column) and disparity map added with Canny edges (right column). (a) Disparity map at the coarsest level ($l = 3$). (b) Disparity map at finer level ($l = 2$).

noted that the proposed adaptive penalties method can obtain accurate disparities.

E. IMPLEMENTATION DETAILS

For traditional SGM and MGM methods, the problems are the computational complexity and matching ambiguities mainly caused by the full disparity search range. The advantage of using 3-D plane fitting and coarse-to-fine disparity constraint is that the pixel-wise disparity search range is restricted to just a few pixels. The benefit of using adaptive penalties is penalizing discontinuities and occlusions for accurate disparity. The overall procedure of our proposed algorithm is summarized in Algorithm 2.

Algorithm 2 Overall Procedure of Our Proposed Method

Input: Two epipolar rectified left and right images.
Output: Disparity maps of the input left and right images.
1. A stereo pair of Gaussian pyramids with L levels are constructed from the input two rectified images.
2. For the two images at the coarsest level $l = L$
   2.1 Extract and match SIFT keypoints, fit 3-D plane, and compute pixel-wise disparity search range based on the fitted 3-D plane.
   2.2 Compute matching cost and aggregation with adaptive penalties.
   2.3 After running the above steps on both left and right images at the coarsest level independently, do post processing including subpixel estimation and left-to-right consistency check.
3. **for** $l = L - 1$ to 1 do
   3.1 Coarse-to-fine disparity constraint: the disparity map at a coarser level $l + 1$ is used to restrict pixel-wise disparity search range at a finer level l.
   3.2 Compute matching cost and aggregation with adaptive penalties.
   3.3 After running the above steps on both left and right images at level l independently, do post processing including subpixel estimation and left-to-right consistency check.
   end for l
4. Finally, when the process comes to the finest level $l = 1$ that is, the input two rectified images, we can get the final left and right subpixel disparity maps.

III. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL SETTINGS

To evaluate the performance of the proposed method, we use real-world stereo image pairs[1] taken by Chang'e-3 Yutu rover panoramic camera (PCAM) system. These stereo images must be rectified firstly by the epipolar rectification method [31], and the sizes of rectified stereo images are 1,176 × 864 pixels.

The left images of the four stereo pairs used in the paper are shown in Fig. 8. The test images contain considerably difficult regions, such as disparities discontinuous, occlusion, low texture or textureless, similar texture, and image noises regions. As shown in Table 1, the disparity ranges of the four stereo pairs are about $[-40, 150]$, $[-97, -9]$, $[-102, 8]$, $[-106, 3]$, respectively. Furthermore, considering the lack of standard datasets with ground truth of unstructured environments for quantitative evaluation of the accuracy of our proposed method, we chose the KITTI dataset[2] which also considers an environment dominated by a slanted plane. The KITTI dataset was created from a driving platform which is similar to the planetary rover in the unstructured environment. This dataset contains real-world complex illumination conditions and large displacements [39].

In experiments, our proposed method are implemented with the following parameter settings. During the coarse-to-fine processing, we construct two 3-level pyramids ($L = 3$) for the rectified left and right images. The parameter used in section II-A is the disparity search offset Δd, which is set to $\Delta d = 2$ for the stereo pairs of Chang'e-3 Yutu rover, $\Delta d = 5$ for the stereo pairs of KITTI dataset, considering the KITTI dataset contains a large portion of tall obstacles, such as buildings and trees. The parameters used in section II-B are disparity threshold ΔT, which is set to $\Delta T = 1.5, 3.0$ and 4.5 respectively for three levels, and the original penalties, which are set to $P_1 = 8$ and $P_2 = 32$. The parameter used in section II-C is the disparity search offset ΔD at a finer level, which is set to $\Delta D = 1$. The method was executed on a desktop PC with i5-8250 1.60Hz CPU and 8-GB memory, and our code are written in Microsoft Visual Studio 2013 with C++ and OpenCV library.

B. PERFORMANCE AT THE COARSEST LEVEL

The disparity map generated at the coarsest level ($l = 3$) is important for the disparity maps generation at other levels.

[1] http://planetary.s3.amazonaws.com/data/change3/pcam.html
[2] http://www.cvlibs.net/datasets/kitti/

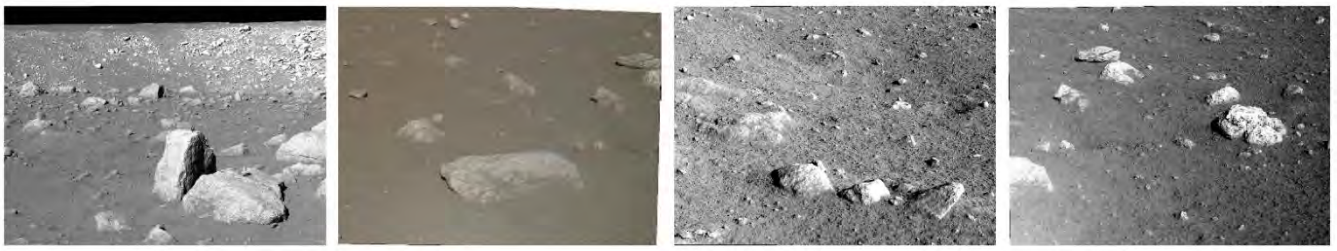

FIGURE 8. The four rectified left images of stereo pair-1, pair-2, pair-3, and pair-4 for Chang'e-3 Yutu rover, with 1, 176 × 864 pixels, respectively from left to right.

TABLE 1. Four stereo image pairs of Chang'e-3 Yutu rover.

Stereo pair	Image size	Disparity range/pixels	Main characteristic
pair-1	1,176 × 864	[-40, 150]	Huge rock, discontinuous, occlusion
pair-2	1,176 × 864	[-97, -9]	Huge rock, discontinuous, occlusion, weak texture, or textureless
pair-3	1,176 × 864	[-102, 8]	Rock, discontinuous, occlusion
pair-4	1,176 × 864	[-106, 3]	Rock, discontinuous, occlusion, weak texture, repetitive patterns, varying light

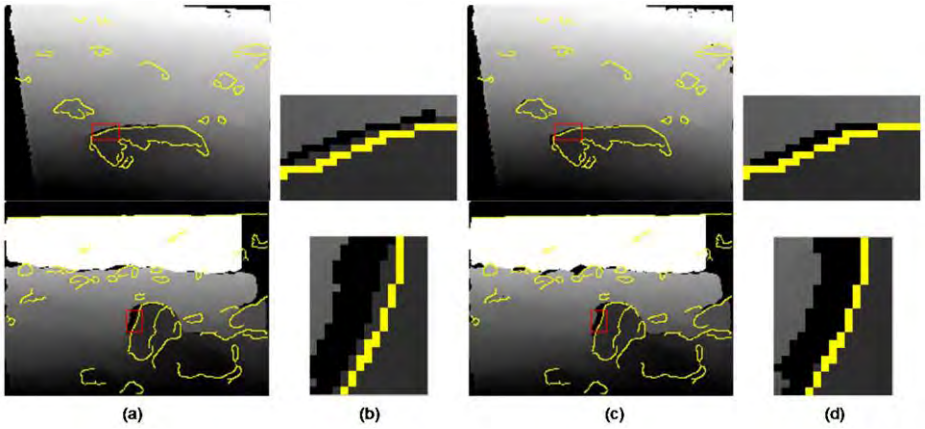

FIGURE 9. Examples of disparity maps at the coarsest level, added with edges. (a) Disparity map based on constant penalties. (b) Zoomed-in disparity map of (a), (c) Disparity map based on adaptive penalties. (d) Zoomed-in disparity map of (c).

To study the effect of penalties at the coarsest level, we evaluate the performance of two different schemes. The first scheme uses constant penalties instead of adaptive penalties in the cost computation and aggregation step. Fig. 9(a) and (b) show the disparity maps generated by the first scheme for pair-1 and pair-2 at the coarsest level, where the Canny edges are marked in yellow color. Note that it fails to obtain correct disparity in disparity discontinuities and occlusion regions. The second scheme uses our proposed 3-step scheme with adaptive penalties at the coarsest level. Fig. 9(c) and (d) show the disparity maps for pair-1 and pair-2 at the coarsest level with the second scheme, and we can find that it has achieved correct disparity at depth discontinuities, edge regions, and occlusion regions.

We further test the processing time with different disparity search offset Δd. As shown in Fig. 10, it spends more time with large disparity search offset, while it is sufficient to yield a good disparity map when Δd=2 (see Fig. 9(c) and (d)).

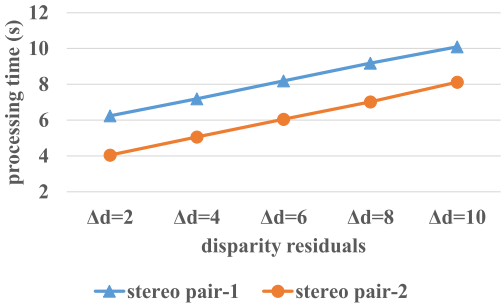

FIGURE 10. Effects of different disparity search offsets for stereo pair-1 and pair-2.

C. PERFORMANCE OF THE PROPOSED METHOD

We validate our proposed method with data captured from lunar surface, and the disparity maps are shown in Fig. 11(a). The common properties of these four stereo pairs are that they all contain rocks, occlusions, repetitive patterns, and low

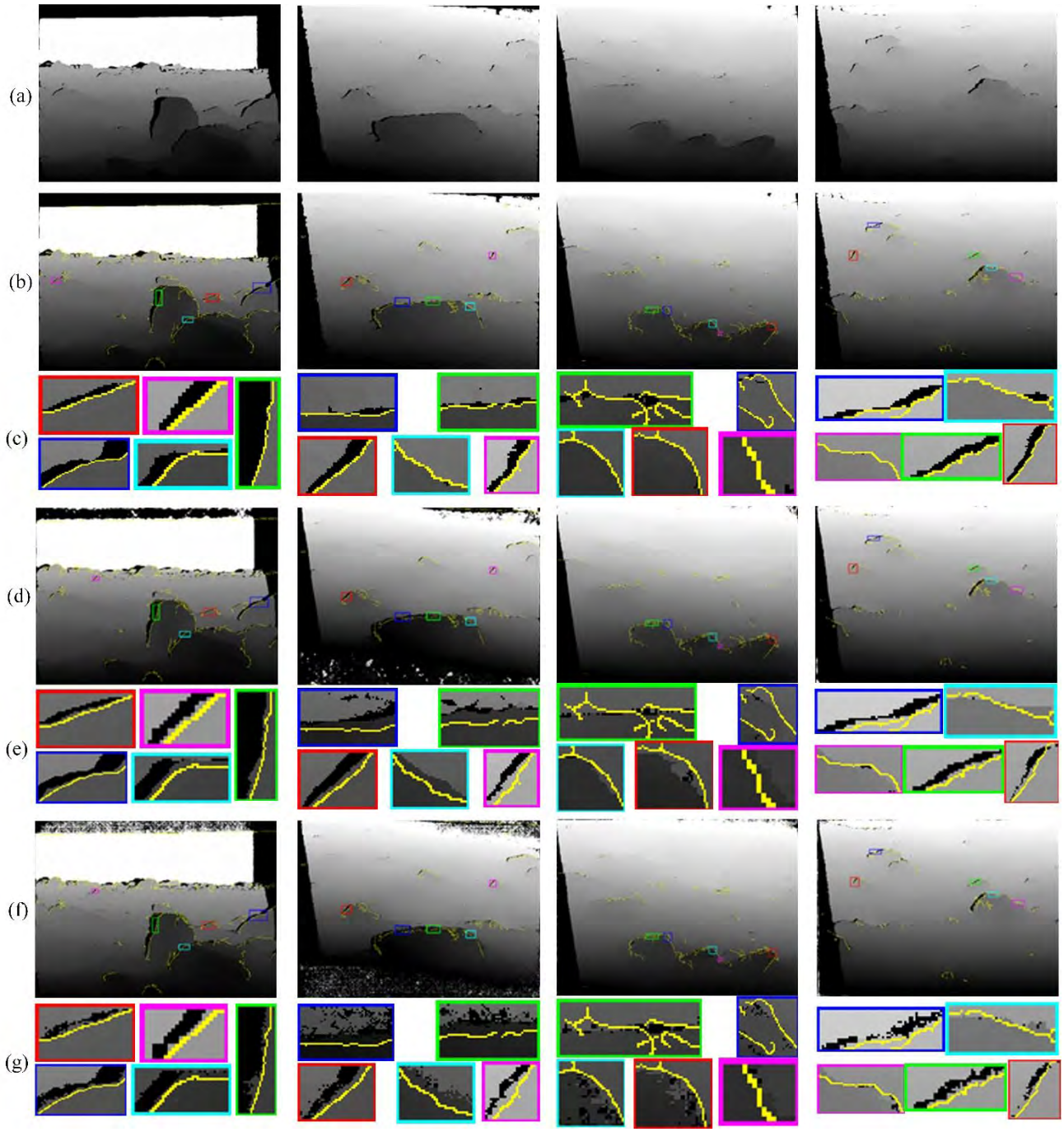

FIGURE 11. Comparison of disparity maps generated by our method, MGM, and SGM. From left to right, they are disparity maps of stereo pair-1, pair-2, pair-3, and pair-4, respectively. Each column shows, from top to bottom, (a) Disparity map estimated by our proposed method. (b) Disparity map estimated by our proposed method (added with edge points). (c) The zoomed-in regions from (b). (d) Disparity map estimated by MGM method (added with edge points). (e) The zoomed-in regions from (d). (f) Disparity map estimated by SGM method (added with edge points). (g) The zoomed-in regions from (f). Note that our proposed method can obtain the correct disparity results even in the discontinuous and occlusion regions, while MGM and SGM methods cannot obtain the correct results in these regions.

texture or even textureless regions. However, these difficulties are very common on planetary surface. For example, the pair-1 has huge rocks, precipice, disparity discontinuous, and occlusion regions, the pair-2 has huge rock, disparity discontinuous and occlusion regions, many low texture or textureless regions, the pair-3 contains rocks, repetitive patterns, and the pair-4 has rocks, varying light conditions. To further evaluate the disparity accuracy of our method, we show disparity maps added with Canny edges (see Fig. 11(b)) which are marked in yellow. We can see that all the rocks on stereo images have been well reconstructed, and the scene behind the cliff on stereo pair-1 is also perfectly reconstructed, shown

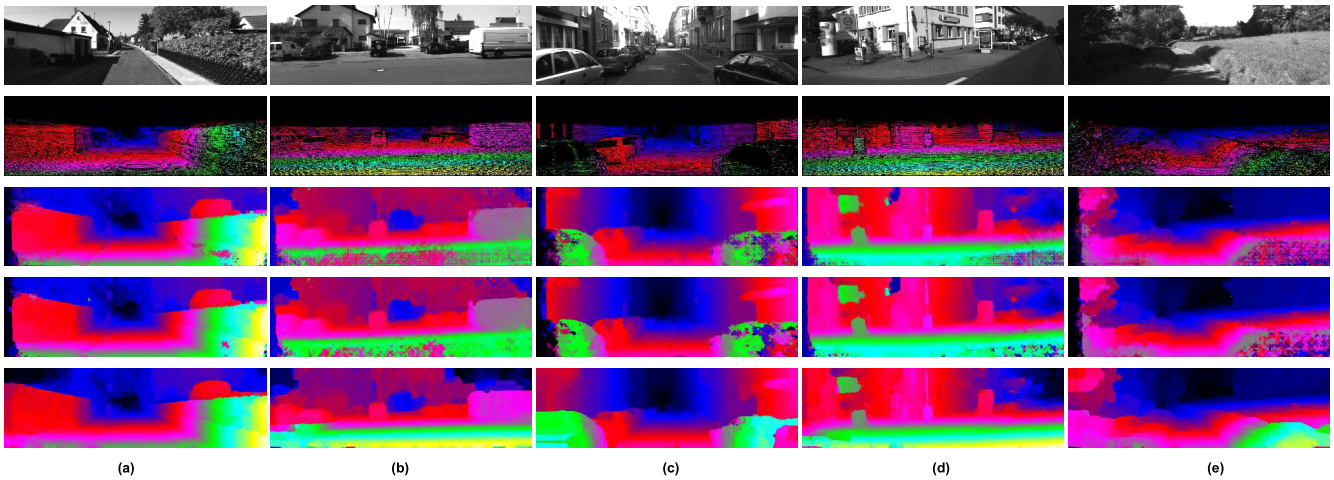

FIGURE 12. Examples of disparity maps of five frames in KITTI training pairs with different methods. Disparity maps are visualized using color-map. (a) From top to bottom: No. 14 left image, ground truth pixels, disparity maps of SGM, MGM, and the proposed method. (b) From top to bottom: No. 69 left image, ground truth pixels, disparity maps of SGM, MGM, and the proposed method. (c) From top to bottom: No. 83 left image, ground truth pixels, disparity maps of SGM, MGM, and the proposed method. (d) From top to bottom: No. 147 left image, ground truth pixels, disparity maps of SGM, MGM, and the proposed method. (e) From top to bottom: No. 183 left image, ground truth pixels, disparity maps of SGM, MGM, and the proposed method.

in white. Specifically, from the disparity map of stereo pair-2 (see the second column of Fig. 11(b)), it is observed that our proposed method is capable of addressing the matching ambiguities at low texture or textureless regions. To get a clearer view, we show five zoomed-in regions added with Canny edges for each disparity map (see Fig. 11(c)). As seen clearly from these zoomed-in regions, our proposed method can obtain accurate disparity especially at disparities discontinuities and occlusion regions.

D. COMPARISON OF DIFFERENT METHODS

We compare our proposed method with two state-of-the-art semi-global matching methods, i.e., MGM and SGM methods, and evaluate the performance of accuracy and efficiency. All the disparity maps of SGM and MGM are generated by the online MGM demo.[3] The MGM and SGM methods use the same following settings: eight propagation directions and census transform for matching cost on a 3×3 neighborhood.

1) ACCURACY ANALYSIS

We test the accuracy of disparity maps generated by SGM, MGM, and the proposed method using KITTI dataset. During our experiments, all the 194 training image pairs with ground truth for reflective regions are used. During our experiments, the evaluation metrics are the same as the KITTI benchmark with an error threshold 3. We show the disparity maps estimated by SGM, MGM, and our proposed method in Fig. 12. In the given images, there are textureless slanted planes (e.g., roads), different obstacles in front of the camera and on both sides of the road. These pairs consist of different types of obstacles, such as vehicles, buildings, trees, crops and so on. We show the performance of different methods in Table 2, which gives the average quantitative analysis results for all the 194 training image pairs in reflective regions. It is noted that our method performs better than SGM and MGM methods. Fig. 12 shows some examples of disparity maps, and Table 3 gives the corresponding quantitative results of SGM, MGM, and the proposed method. We can see that there are significant improvements in both percentage of erroneous pixels and average disparity error with our proposed method. For low texture or textureless areas, e.g., roads in Fig. 12(b) and (d), and repetitive texture areas, e.g., crops in Fig. 12(e), SGM and MGM methods tend to generate errors due to large disparity search range. While in disparity map generated by our proposed method, the errors are almost eliminated at these regions.

Further, we compare our method with SGM and MGM methods using the real-world image pairs obtained by Chang'e-3 Yutu rover. As shown in Fig. 11, the disparity map generated by our proposed method is superior to that generated by SGM and MGM methods, and there are some interesting points worth noting. Firstly, the disparity maps generated

TABLE 2. Average quantitative results of SGM, MGM, and the proposed method for the training images on KITTI dataset using the default error threshold of 3 pixels in reflective regions. Out-noc: percentage of erroneous pixels in nonoccluded areas. Out-all: percentage of erroneous pixels in pixels in total. Avg-noc: average disparity error in nonoccluded areas. Avg-all: average disparity error in total.

Method	Out-Noc	Out-All	Avg-Noc	Avg-All
SGM	20.63%	22.43%	3.02px	4.13px
MGM	18.05%	19.92%	2.58px	3.73px
Proposed	9.33%	11.25%	1.59px	1.84px

[3] http://dev.ipol.im/∼ facciolo/mgm/

TABLE 3. Quantitative results of SGM, MGM, and the proposed method for No. 14, No. 69, No. 83, No. 147, and No. 183 on KITTI dataset using the default error threshold of 3 pixels in reflective regions.

Image No.	SGM				MGM				Proposed method			
	Out-Noc	Out-All	Avg-Noc	Avg-All	Out-Noc	Out-All	Avg-Noc	Avg-All	Out-Noc	Out-All	Avg-Noc	Avg-All
No. 14	14.27%	16.49%	2.07px	3.00px	11.01%	13.31%	1.67px	2.63px	4.44%	6.67%	0.83px	0.97px
No. 69	46.2%	47.1%	7.91px	8.27px	42.33%	43.29%	6.02px	6.43px	10.18%	11.31%	1.25px	1.32px
No. 83	19.47%	20.46%	3.42px	4.38px	16.72%	17.74%	3.09px	4.05px	10.75%	11.85%	1.90px	2.10px
No. 147	28.04%	29.30%	4.38px	4.80px	25.16%	26.48%	3.42px	3.90px	3.92%	5.14%	0.73px	0.81px
No. 183	26.40%	27.79%	6.58px	7.65px	23.01%	24.46%	5.61px	6.74px	10.73 %	12.42 %	1.60px	1.97px

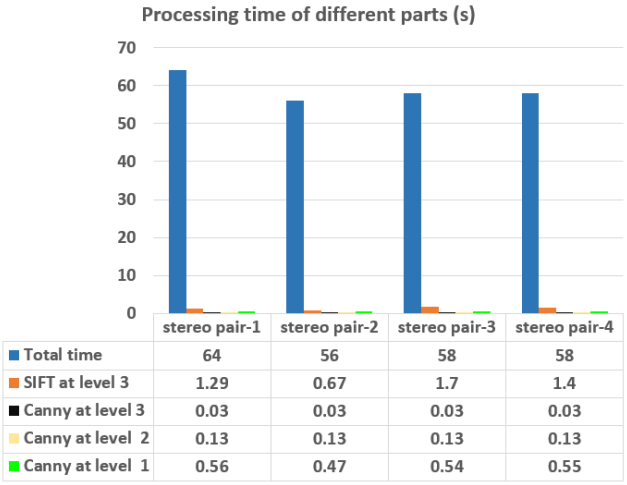

FIGURE 13. Processing time of the total and additional parts for the four stereo pairs.

by MGM and SGM methods are often incorrect at depth discontinuities, edge areas, and the occlusion regions (see Fig. 11(e) and (g)), however, our method can produce correct disparities in these regions (see Fig. 11(c)). The main reason is that our method uses adaptive penalties based on Canny edges in matching cost aggregation. Secondly, as shown at the bottom of disparity map of stereo pair-2 generated by SGM and MGM (see the second column of Fig. 11(d) and (f)), there are errors in textureless lunar surfaces. While for the disparity map generated by our method, errors in the textureless areas are fully addressed (see the second column of Fig. 11(b)). This is because the proposed method has small disparity search ranges, while SGM and MGM have large disparity search ranges which may result in errors especially at low texture or textureless regions.

2) EFFICIENCY ANALYSIS

We approximately compare the computational complexity of the proposed method with that of MGM method for the four stereo pairs. Compared with MGM method, our proposed method adds the following additional processing parts, i.e., SIFT keypoints extraction and matching at the coarsest level, and Canny edge detection at all levels. Fig. 13 shows the processing time of the total and additional parts in 3 levels pyramid for the four stereo pairs. We can see that the processing time of SIFT extraction and matching at the coarsest level ($l = 3$) only accounts for a very small part of the total time, while processing time of Canny edge detection at all levels accounts for less. Since the computational complexity analysis is only an approximate estimation, the computational complexity of SIFT and Canny is not considered. Suppose the image size is $W \times H$ pixels and the fixed disparity range is D pixels, therefore, the computational complexity of MGM is given by (12).

$$O(T_{\mathrm{MGM}}) = O(W \times H \times D) \tag{12}$$

However, disparity search range of our proposed method is adaptively pixel-wise determined, therefore, the computational complexity of our method is estimated by the sum of disparity range for all the pixels as shown in (13).

$$O(T_{\mathrm{Pr}\,oposed}) = \sum_{j=1}^{H} \sum_{i=1}^{W} O(d_{ij}) \tag{13}$$

where d_{ij} is the disparity range of pixel (i, j), $d_{ij} = D_{\max}(i, j) - D_{\min}(i, j) + 1$.

Table 4 gives the approximately computational complexity of MGM method estimated by (12) and that of our method estimated by (13), where O(L-R) is the complexity of left-to-right stereo matching, and O(R-L) is the complexity of right-to-left stereo matching. The computational complexity of our proposed method is obtained by adding the complexity of all three levels. Finally, the ratio between the complexity of our method and that of MGM can be computed. As shown in Table 4, our proposed method can greatly reduce the computational complexity for all the test images, which is only a few percent of that of MGM method.

Furthermore, we compare the processing time of the proposed method against that of MGM method based on the four stereo pairs of Yutu rover, and the results are shown in Fig. 14. It should be noted that for fair matching comparison, the processing time of MGM was obtained by running the downloaded C++ source code[4] of MGM on our desktop PC (described in section III-A). It is quite obvious that our proposed method greatly reduces the processing time and improves the processing efficiency. For example, the generation of disparity map for stereo pair-1 with MGM takes about 1,008 seconds, which takes the longest time because its full disparity search range $[-40, 150]$ is the largest. While it takes about 486 seconds for stereo pair-2 with MGM, which takes the shortest time because its full disparity range $[-97, -9]$ is the shortest. However, the processing time with our method is almost the same for the four stereo pairs, because the disparity

[4]https://github.com/gfacciol/mgm

TABLE 4. Approximate comparison of computational complexity between our proposed method and MGM.

Stereo Pair		Our proposed method			Total complexity of our proposed method	MGM	Total complexity of MGM	The ratio of total complexity (%)
		Level-3	Level-2	Level-1				
1	O(L-R)	O(1,194,989)	O(1,182,535)	O(4,635,320)	O(14,171,402)	O(193,052,160)	O(386,104,320)	3.67%
	O(R-L)	O(1,194,310)	O(1,183,284)	O(4,780,964)		O(193,052,160)		
2	O(L-R)	O(763,692)	O(967,525)	O(4,188,161)	O(11,609,077)	O(89,413,632)	O(178,827,264)	6.49%
	O(R-L)	O(750,941)	O(970,509)	O(3,968,249)		O(89,413,632)		
3	O(L-R)	O(712,475)	O(952,224)	O(4,134,032)	O(11,349,885)	O(111,767,040)	O(223,534,080)	5.07%
	O(R-L)	O(708,146)	O(949,692)	O(3,893,316)		O(111,767,040)		
4	O(L-R)	O(831,310)	O(959,788)	O(4,176,512)	O(11,684,929)	O(110,750,976)	O(221,501,952)	5.27%
	O(R-L)	O(839,799)	O(957,116)	O(3,920,404)		O(110,750,976)		

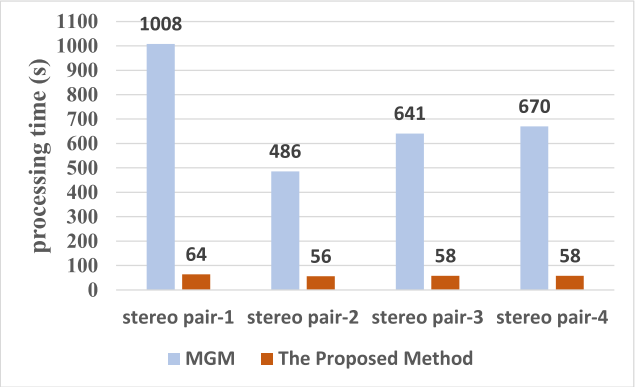

FIGURE 14. Processing time for the four stereo pairs of Chang'e-3 Yutu rover based on MGM method and our proposed method.

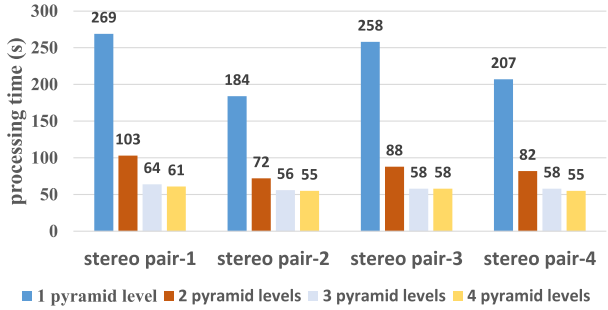

FIGURE 15. Processing time with different numbers of pyramid levels for the four stereo pairs.

search range of our method is pixel-wise determined, which is not affected by the full disparity search range of the whole image.

In the proposed method, for the disparity map generation at the coarsest level, we propose 3-D plane fitting to restrict disparity search range to a few pixels. For disparity map generation at a finer level, the pixel-wise disparity search range is restricted to several pixels based on the obtained disparity map at coarser level. While MGM method is performed with full disparity search range, which causes the execution of algorithm less efficient.

E. PARAMETERS INFLUENCE ANALYSIS

To get a better analysis of our method, we evaluate the influences of different parameter settings. After obtaining the results using the parameters set in section III-A, we can systematically vary the main parameters and obtain the disparity results.

The first key parameter is the number of pyramid levels L. If this parameter is set too small, it will be difficult to improve the stereo matching speed. If it is too large, the image size at the coarsest level will be very small. Firstly, the lower the image resolution, the fewer scene details will be contained. Therefore, the number of SIFT keypoints at the coarsest level is often too small to fit a 3-D plane. Secondly, it has little effect on improving the processing speed, even if the number of pyramid levels is larger. In Fig. 15, we vary L from 1 to 4 and obtain the processing time of our method based on the four stereo pairs. We can see that it consumes the most time when $L = 1$, i.e., without pyramid decomposition. As the number of pyramid levels increases, the time spend is rapidly decreasing. However, when $L \geq 3$, the change of processing time is not obvious. We have also tested the matching results when $L = 5$, however, few or no matched SIFT keypoints can be obtained at the coarsest level due to the lower image resolution. Furthermore, Fig. 16 shows the disparity maps with pyramid levels from 1 to 4 for the four stereo pairs of Yutu rover. It can be seen that when $L = 1$, the quality of disparity map is the worst, especially in the areas with large and prominent rocks. Therefore, for our test images with $1{,}176 \times 864$ pixels, when the number of pyramid levels is set to 3, we get a high quality disparity map, and meanwhile it has the lowest computational cost.

Another parameter influencing the running speed and accuracy is the disparity offset Δd in 3-D plane fitting at the coarsest level. If this parameter is set too large, the disparity search range becomes large, causing more computational overhead. However, if Δd is set too small, disparities of the obstacles at the coarsest level are not within the disparity search range. We set to $\Delta d = 2$ for Yutu rover images, and $\Delta d = 5$ for the KITTI dataset because of the tall buildings and trees existed in this dataset.

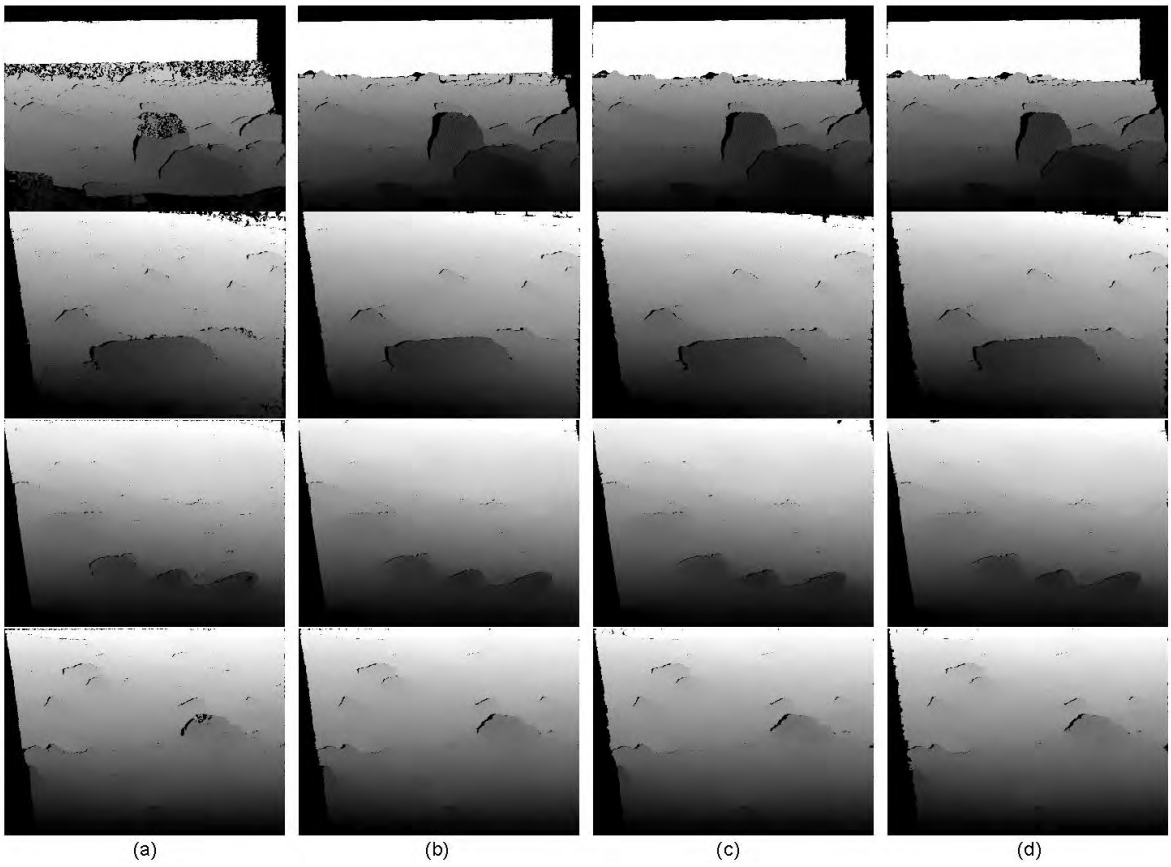

FIGURE 16. The influences of different numbers of pyramid levels for the four pairs: (a) 1 level, (b) 2 levels, (c) 3 levels, (d) 4 levels. Each column shows, from top to bottom, the disparity maps of pair-1, pair-2, pair-3, and pair-4, with the same number of pyramid level. We can see that when the number of pyramid levels is 1, the disparity map is the worst, especially in the regions containing large and prominent rocks.

IV. CONCLUSION

To generate disparity map accurately and efficiently from stereo images captured by a planetary rover in unstructured environment of the planetary surface, we propose a coarse-to-fine hierarchal method for automatic dense stereo matching. The method starts from the pyramid construction of two input rectified images. At the coarsest level, the stereo matching for the two coarsest images is implemented by a 3-step scheme, i.e., (1) 3-D plane fitting, (2) cost computation and aggregation with adaptive penalties, and (3) post processing, which serves as a good starting point for processing at the following finer levels. The proposed 3-D plane fitting method is utilized to adaptively determine pixel-wise disparity search range to reduce the computation cost. Our proposed adaptive penalties method can improve the matching accuracy at the depth discontinuities, edge areas, and the occlusion regions. At finer levels, the stereo matching is implemented by another 3-step scheme without 3-D plane fitting, which uses coarse-to-fine disparity constraint instead of the 3-D plane fitting constraint to greatly reduce the disparity search range, i.e., (1) coarse-to-fine disparity constraint, (2) cost computation and aggregation with adaptive penalties, and (3) post processing. We also conduct the qualitative and quantitative evaluations of our matching method which demonstrate the accuracy and efficiency of the proposed method in real-world scenes. The generated disparity maps can enable rover safe, power-efficient automatic motion in unstructured environment of the planetary surface. This method was originally developed for the applications of planetary rover, but can also be used for other applications, such as field robot in an environment dominated by a slanted plane.

ACKNOWLEDGMENT

The authors would like to thank the editors and anonymous reviews for their valuable comments and helpful suggestions which greatly improve the paper's quality.

REFERENCES

[1] E. Krotkov and R. Hoffman, "Terrain mapping for a walking planetary rover," *IEEE Trans. Robot. Autom.*, vol. 10, no. 6, pp. 728–739, Dec. 1994.
[2] S. Se, H. K. Ng, P. Jasiobedzki, and T. J. Moyung, "Vision based modeling and localization for planetary exploration rovers," in *Proc. 55th Int. Astron. Congr.*, 2004, pp. 434–440.
[3] S. B. Goldberg, M. Maimone, and L. Matthies, "Stereo vision and rover navigation software for planetary exploration," in *Proc. IEEE Aerosp. Conf.*, vol. 5, Mar. 2002, pp. 5-2025–5-2036.
[4] L. Matthies, T. Balch, and B. Wilcox, "Fast optical hazard detection for planetary rovers using multiple spot laser triangulation," in *Proc. IEEE Int. Conf. Robot. Autom.*, vol. 1, Apr. 1997, pp. 859–866.

[5] G. N. DeSouza and A. C. Kak, "Vision for mobile robot navigation: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 2, pp. 237–267, Feb. 2002.
[6] G. R. Deen and J. J. Lorre, "Seeing in three dimensions: Correlation and triangulation of Mars Exploration Rover imagery," in *Proc. IEEE Int. Conf. Syst., Man Cybern.*, vol. 1, Oct. 2005, pp. 911–916.
[7] J. F. Bell *et al.*, "The Mars Science Laboratory *Curiosity* rover Mastcam instruments: Preflight and in-flight calibration, validation, and data archiving," *Earth Space Sci.*, vol. 4, pp. 396–452, Jul. 2017.
[8] S. Se, T. Barfoot, and P. Jasiobedzki, "Visual motion estimation and terrain modeling for planetary rovers," in *Proc. 8th Int. Symp. Artif. Intell. Robot. Automat. Space*, 2005, pp. 5–8.
[9] J. Yang, H. Wang, Z. Ding, Z. Lv, W. Wei, and H. Song, "Local stereo matching based on support weight with motion flow for dynamic scene," *IEEE Access*, vol. 4, pp. 4840–4847, 2016.
[10] M. Vergauwen, M. Pollefeys, and L. Van Gool, "A stereo-vision system for support of planetary surface exploration," *Mach. Vis. Appl.*, vol. 14, no. 1, pp. 5–14, Apr. 2003.
[11] E. Krotkov and M. Hebert, "Mapping and positioning for a prototype lunar rover," in *Proc. IEEE Int. Conf. Robot. Automat.*, May 1995, pp. 2913–2919.
[12] A. Buades and G. Facciolo, "Reliable multiscale and multiwindow stereo matching," *SIAM J. Imag. Sci.*, vol. 8, no. 2, pp. 888–915, 2015.
[13] S. Li, K. Chen, M. Song, D. Tao, G. Chen, and C. Chen, "Robust, efficient depth reconstruction with hierarchical confidence-based matching," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3331–3343, Jul. 2017.
[14] N. Komodakis, G. Tziritas, and N. Paragios, "Fast, approximately optimal solutions for single and dynamic MRFS," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2007, pp. 1–8.
[15] 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.
[16] 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.
[17] A. F. Bobick and S. S. Intille, "Large occlusion stereo," *Int. J. Comput. Vis.*, vol. 33, no. 3, pp. 181–200, 1999.
[18] A. Klaus, M. Sormann, and K. Karner, "Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure," in *Proc. Int. Conf. Pattern Recognit*, 2006, pp. 15–18.
[19] M. Peng, Y. Liu, Z. Liu, and K. Di, "Global image matching based on feature point constrained Markov Random Field model for planetary mapping," in *Proc. 32nd Asian Conf. Remote Sens.*, 2011, pp. 373–378.
[20] Y.-C. Wang, C.-P. Tung, and P.-C. Chung, "Efficient disparity estimation using hierarchical bilateral disparity structure based graph cut algorithm with a foreground boundary refinement mechanism," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 5, pp. 784–801, May 2013.
[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.
[22] K. R. Kim, Y. J. Koh, and C. S. Kim, "Multiscale feature extractors for stereo matching cost computation," *IEEE Access*, vol. 6, pp. 27971–27983, 2018.
[23] R. Barnes *et al.*, "Geological analysis of martian rover-derived digital outcrop models using the 3-D visualization tool, planetary robotics 3-D viewer—PRo3D," *Earth Space Sci.*, vol. 5, pp. 285–307, Nov. 2018. [Online]. Available: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2018EA000374
[24] M. Balme *et al.*, "Surface-based 3D measurements of small aeolian bedforms on Mars and implications for estimating ExoMars rover traversability hazards," *Planet. Space Sci.*, vol. 153, pp. 39–53, Apr. 2018.
[25] M. Rothermel, K. Wenzel, and D. Fritsch, "SURE: Photogrammetric surface reconstruction from imagery," in *Proc. LC3D Workshop*, 2012, pp. 1–21.
[26] S. Hermann and R. Klette, "Evaluation of a new coarse-to-fine strategy for fast semi-global stereo matching," in *Proc. 5th Pacific-Rim Symp. Image Video Technol. (PSIVT)* (Lecture Notes in Computer Science), vol. 7087, 2012, pp. 395–406. [Online]. Available: https://cerv.aut.ac.nz/wp-publications/hermann2011evaluationmatching/
[27] D. Hernandez-Juarez, A. Chacón, A. Espinosa, D. Vázquez, J. C. Moure, and A. M. López, "Embedded real-time stereo estimation via semi-global matching on the GPU," *Procedia Comput. Sci.*, vol. 80, pp. 143–153, Dec. 2016.
[28] C. Banz, S. Hesselbarth, H. Flatt, H. Blume, and P. Pirsch, "Real-time stereo vision system using semi-global matching disparity estimation: Architecture and FPGA-implementation," in *Proc. Int. Conf. Embedded Comput. Syst., Archit., Modeling Simulation*, Jul. 2010, pp. 93–101.
[29] F. Qamar, F. B. Muslim, F. Gregoretti, L. Lavagno, and M. T. Lazarescu, "High-level synthesis for semi-global matching: Is the juice worth the squeeze?" *IEEE Access*, vol. 5, pp. 8419–8432, 2017.
[30] G. Facciolo, C. De Franchis, and E. Meinhardt, "MGM: A significantly more global matching for stereovision," in *Proc. Brit. Mach. Vis. Conf.*, 2015, pp. 90.1–90.12. [Online]. Available: http://www.bmva.org/bmvc/2015/papers/paper090/index.html
[31] P. Monasse, "Quasi-Euclidean epipolar rectification," *Image Process. Line*, vol. 1, pp. 187–199, Sep. 2011. [Online]. Available: http://www.ipol.im/pub/art/2011/m_qer/
[32] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. COM-31, no. 4, pp. 532–540, Apr. 1983.
[33] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[34] M. A. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
[35] G. A. Kordelas, D. S. Alexiadis, P. Daras, and E. Izquierdo, "Content-based guided image filtering, weighted semi-global optimization, and efficient disparity refinement for fast and accurate disparity estimation," *IEEE Trans. Multimedia*, vol. 18, no. 2, pp. 155–170, Feb. 2016.
[36] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, Nov. 1986.
[37] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Proc. Eur. Conf. Comput. Vis.*, vol. 801, 1994, pp. 151–158.
[38] I. Haller and S. Nedevschi, "Design of interpolation functions for subpixel-accuracy stereo-vision systems," *IEEE Trans. Image Proc.*, vol. 21, no. 2, pp. 889–898, Feb. 2012.
[39] R. Urtasun, P. Lenz, and A. Geiger, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3354–3361.

HAICHAO LI received the B.S. and M.S. degrees in mechanical and electrical engineering from the Beijing University of Chemical Technology, Beijing, China, in 2001 and 2004, respectively, and the Ph.D. degree from the School of Instrument Science and Optoelectronics Engineering, Beijing University of Aeronautics and Astronautics, Beijing, in 2008. He is currently an Associate Professor with the Qian Xuesen Laboratory of Space Technology, China Academy of Space Technology. His research interests include stereo vision, 3-D reconstruction, motion detection and tracking, and deep learning.

LIANG CHEN received the B.S. and M.S. degrees in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2003 and 2006, respectively, and the Ph.D. degree from the State Key Laboratory of Remote Sensing Science, Chinese Academic of Sciences, Beijing, China, in 2009. He is currently a Senior Engineer with the Qian Xuesen Laboratory of Space Technology, China Academy of Space Technology. His research interests include image processing, image registration, data mining, and parameters retrieval.

FENG LI (M'07–SM'10) received the B.S.E.E. degree from Lanzhou Railway University, in 1999, the M.Eng. degree from the China Academy of Space Technology, in 2002, and the Ph.D. degree in electrical engineering from The University of New South Wales, in 2009. He was involved in astronomical image processing with CSIRO, Australia, and remote sensing image processing with the Chinese Academy of Science. He is currently the PI of the Qian Xuesen Laboratory of Space Technology, China Academy of Space Technology. His research interests include image registration, super resolution, and compressive sensing.