IEEE Robotics and Automation Letters (RAL) paper presented at the
2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 25-29, 2020, Las Vegas, NV, USA (Virtual)

Virtual IR Sensing for Planetary Rovers: Improved Terrain Classification and Thermal Inertia Estimation

Yumi Iwashita[1], Kazuto Nakashima[2], Joseph Gatto[3], Shoya Higa[1], Adrian Stoica[1], Norris Khoo[3], Ryo Kurazume[2]

***Abstract*— Terrain classification is critically important for Mars rovers, which rely on it for planning and autonomous navigation. On-board terrain classification using visual information has limitations, and is sensitive to illumination conditions. Classification can be improved if one fuses visual imagery with additional infrared (IR) imagery of the scene, yet unfortunately there are no IR image sensors on the current Mars rovers. A virtual IR sensor, estimating IR from RGB imagery using deep learning, was proposed in the context of a MU-Net architecture. However, virtual IR estimation was limited by the fact that slope angle variations induce temperature differences within the same terrain. This paper removes this limitation, giving good IR estimates and as a consequence improving terrain classification by including the additional angle from the surface normal to the Sun and the measurement of solar radiation. The estimates are also useful when estimating thermal inertia, which can enhance slip prediction and small rock density estimation. Our approach is demonstrated in two applications. We collected a new data set to verify the effectiveness of the proposed approach and show its benefit by applying to the two applications.**

Index Terms— Space Robotics and Automation, Multi-Modal Perception

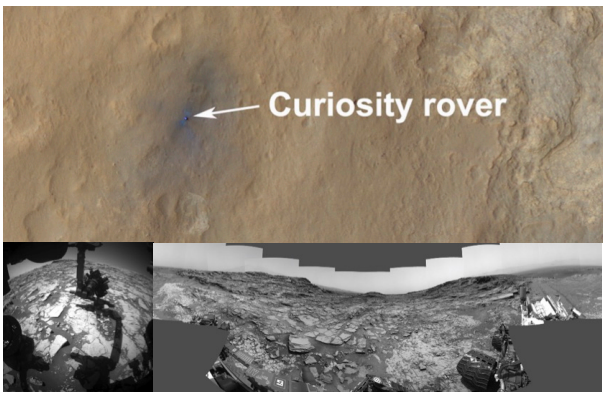

Fig. 1. Top Image from High Resolution Imaging Science Experiment (HiRISE) camera on board the Mars Reconnaissance Orbiter has a 0.3m/pixel resolution, resolving objects below a meter across; bottom left image shows also the Curiosity rover; bottom right image from front Hazard avoidance camera and from 360-degree panorama from the Navigation Camera

I. Introduction

The primary information sources used to evaluate terrain traversibility and conduct surface mission planning for planetary rover missions are remote imaging sensors on-board satellites and image sensors on-board planetary rovers. While executing mission plans, a rover must rely on autonomous navigation, for which the rover is dependent on its on-board sensors. The Mars Science Laboratory (MSL), 'Curiosity' rover is equipped with multiple sensors including a Navigation Camera (Navcam), Mastcam, front and rear Hazard Avoidance Cameras (Hazcams), etc., for both science and engineering objectives. Figure 1 shows images taken by the HiRISE camera on board the Mars Reconnaise Orbiter and from the front Hazcam and Navcam on-board the MSL rover. Power constraints limit the use of sensors for MSL driving; and while the radioisotope thermoelectric generator (RTG) is able to provide a constant level of power, the majority of the power is used for mobility of the heavy rover.

Hazcams and Navcams are grayscale cameras covering red wavelengths centered at $\sim$650 nanometers. The Mast Camera, or Mastcam for short, takes color images and color video footage of the Martian terrain. Visible domain imagery is used to perform terrain classification [1], yet it has sensitivity to illumination conditions. To increase robustness to illumination changes, the combination of visible and IR imagery has been shown to be more robust [2].

Unfortunately, neither MSL nor the soon to be launched Mars 2020 rover (planned for launch on July 17, 2020) have IR cameras. This could be addressed by the use of a virtual IR sensor, which gives IR estimates based only on RGB input data after training a model using images seen both in RGB and IR. The idea of the virtual IR was proposed in the context of a new neural network architecture called MU-Net (Multiple U-Net) [3]. MU-Net was shown to derive better RGB-to-IR mapping models, however its accuracy was insufficient as the slope angle variations induced temperature differences within the same terrain, making learning difficult.

To overcome this limitation, we consider including additional information, such as the angle from the surface normal to the Sun and the measurement of solar radiation to improve performance. If successful, this approach would be very beneficial for various applications. First, it might improve the performance of terrain classification. Second, one could estimate thermal inertia (defined as the physical property of a material that represents its resistance to changes in temperature) from daytime data. Thermal inertia estimation was proposed for slip prediction [4] and the estimation of small rock density [5].

The main contributions of this paper are the following:

[1]Y. Iwashita, S. Higa, and A. Stoica are with Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr, Pasadena, CA 91109, USA. yumi.iwashita@jpl.nasa.gov

[2]K. Nakashima and R. Kurazume are with Kyushu University, Japan.

[3]J. Gatto is with Columbia University, USA, and N. Khoo is with University Southern California, USA. Both are affiliates with Jet Propulsion Laboratory, California Institute of Technology.

- Propose an improved virtual IR sensor by utilizing the angle from the surface normal to the Sun and the measurement of solar radiation.
- Show the benefit of the improved virtual sensor by applying it to 2 applications: (1) improved terrain classification and (2) thermal inertia estimation. It is important to note that once the improved virtual sensor is trained, both applications no longer need an IR sensor.
- Verify the improved virtual IR sensor and two applications (i.e. terrain classification and thermal inertia estimation) with newly collected data.

To the best of our knowledge, the improved virtual IR sensor is the first approach which utilizes the angle from surface normal to the Sun and the measurement of solar radiation.

The rest of the paper is organized as follows. Section II reviews existing methods for cross-modal deep learning, terrain classification, and thermal inertia estimation. Sections III and IV describe the details of the improved virtual IR sensor within two applications. Section V explains the data set collected and experimental evaluations of the proposed methods with the new data set. Finally, Section VI presents the conclusions and discusses future work.

II. Related works

A. Cross-modal deep learning

The virtual IR sensor is categorized as a cross-modal deep learning task [6] [7] [8] [9]. In [8], Xu proposed a two step approach. First, the Region Reconstruction Network (RRN), which learns a non-linear feature mapping between RGB and thermal image pairs. In the second step, the trained model is transferred into another neural network, whose input and output are RGB image and pedestrian detection results, respectively. Like our virtual sensor network, once the model for RGB-to-IR is trained, the test data set does not require IR information. There are other approaches using GANs (Generative Adversarial Networks) for cross-modal deep learning using [9] [10]. These approaches focus on improving the accuracy of mapping from RGB to IR images. While one can take advantage of the network architectures presented in these approaches, this paper will focus on improving accuracy of the virtual IR sensor by integrating physics and geological information.

B. Terrain classification

Terrain classification is performed by a variety of sensors such as cameras in the visual domain, infrared cameras, LIDAR etc. [11] [12] [13]. Rothrock et al. [1] utilized DeepLab [14] for terrain classification with RGB images from both remote sensing and on-board sensors, and showed its feasibility in terrain classification. There are some existing works which utilize RGB and infrared images, such as [11] [15]. These methods focus on semantic segmentation and classify areas whose interclass variations are relatively huge, such as roads, sky, trees etc. We have proposed TU-Net (Two U-Net) and TDeepLab (Two DeepLab) [2], which combine RGB and IR images efficiently. In this work we showed that the combination of RGB and IR images makes the terrain classifier robust to illumination changes compared with RGB images only. However, if rovers have RGB cameras but no IR camera, the approaches [11] [15] [2] do not work. In this study, however, we show that one can obtain the benefits of utilizing the IR modality in terrain classification with a virtual IR sensor.

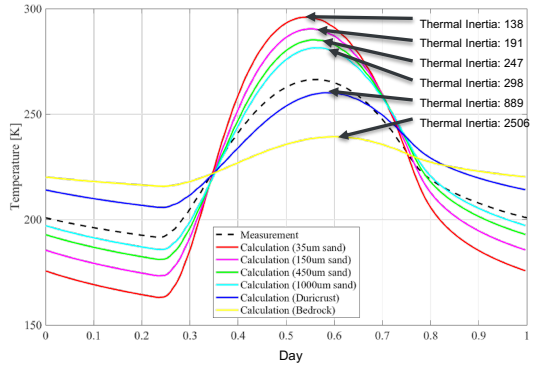

Fig. 2. Thermal inertia examples of 6 material types (35 μm sand, 150 μm sand, 450 μm sand, 1000 μm sand, duricrust, and bedrock).

C. Thermal inertia estimation

Thermal inertia is used in various applications, such as improving rover traversibility [4] [16] and estimating the density of small rocks [5]. Thermal inertia is the physical property of a material highlighting its resistance to changes in temperature. Figure 2 shows how 6 different types of materials (35 μm sand, 150 μm sand, 450 μm sand, 1000 μm sand, duricrust, and bedrock) respond to temperature changes over time. For example, thermal inertia of 35 μm sand with 2.0 $[cm]$ skin depth is 138 $[J^{-2}K^{-1}s^{-1/2}]$, and the one of bedrock with 20.1 $[cm]$ skin depth is much higher 2506 $[J^{-2}K^{-1}s^{-1/2}]$. Thermal inertia can be indirectly estimated by fitting an analytical model to observations of surface temperature [17]. The thermal inertia estimation also requires a thermal sensor, thus rovers without IR cameras cannot estimate thermal inertia.

III. Virtual sensor

In this section we describe the proposed virtual sensor. Critical to the success of the virtual sensor is accounting for variation in thermal information, even within the same terrain type. This occurs because slope angle variations induce temperature differences. Moreover, thermal information changes with time due to the change of solar radiation and angle of the Sun. In order to account for geological information and Sun angle, the following are required: (i) calculation of stereo images and angles to the Sun at each pixel, and (ii) a new deep learning method to efficiently integrate RGB information, angles, and solar radiation.

As for (i), our system has two RGB cameras, an IR camera, a pyranometer (short radiation), and a pyrgeometer (long radiation) (for the system setting, please see more details in experiments). After stereo information is obtained, angle information to the Sun at each pixel is calculated from measured time and location information.

In our previous study [3], we utilized U-Net [18] to learn a RGB to IR mapping. Currently, there are more advanced

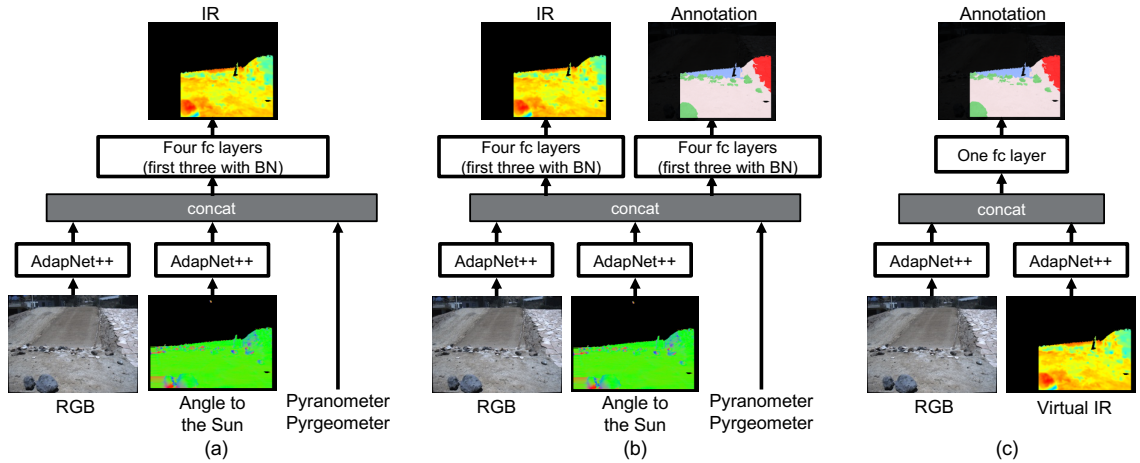

Fig. 3. (a) Network architecture of the proposed Multi-AdapNet++ whose output is thermal information, (b) the architecture whose output is thermal and annotation information, and (c) the architecture for terrain classification

neural network architectures, such as DeepLab [14] and AdapNet++ [19]. In this paper we propose a new network architecture, named Multi-AdapNet++, based on AdapNet++ to efficiently fuse multiple sensor inputs. Figure 3 (a) shows an overview of the proposed Multi-AdapNet++ architecture.

As shown in Fig. 3 (a), we fuse all sensor information via late fusion by concatenating the outputs from the RGB AdapNet++, angle image Adapnet++, and solar radiation data. We note that late fusion was selected after evaluating fusion performance at the early, middle, and late stage. Here, early fusion means that we first concatenate all sensor information, followed by input to AdapNet++. In middle fusion, each modality has its own encoder, and then encoder outputs are concatenated, followed by a decoder. Our experimental results show that the late fusion gives the best performance.

In Fig. 3 (a), the output of the network is an IR image. The loss function $\mathscr{L}_{MSE}$ is based on the MSE (mean squared error),

$$\mathscr{L}_{MSE} = \frac{1}{|S| \times C} \sum_{i \in S} \sum_{j=1}^{C} (a_{ij} - b_{ij})^2, \tag{1}$$

where C is the number of channels, and a_{ij} and b_{ij} are the thermal value at each pixel (i, j) of a ground-truth thermal image a and an output thermal image b, respectively. The loss function is minimized by a stochastic gradient descent method.

In [3], we showed that the use of both thermal and annotated images improves performance. Therefore we utilize two loss functions $\mathscr{L}_{MSE}$ and $\mathscr{L}_{CE}$, (where $\mathscr{L}_{CE}$ refers to a cross entropy loss), as shown in Fig. 3 (b), as

$$\mathscr{L} = \lambda \mathscr{L}_{CE} + \mathscr{L}_{MSE}. \tag{2}$$

Here, λ weighs the impact of the cross entropy loss, which in our experiments was empirically assigned the value 0.20. $\mathscr{L}_{CE}$ is defined as

$$\mathscr{L}_{CE} = -\frac{1}{|S|} \sum_{i \in S} \sum_{j=1}^{N} y_{ij} \log p_{ij}, \tag{3}$$

where N, $|S|$, y_{ij}, p_{ij} are the number of classes, the total number of pixels over the set of images S, ground-truth distribution at each pixel, and outputted probability distribution at each pixel, respectively.

IV. Applications of Virtual Sensor

In this section we highlight two applications of the virtual sensor. These applications no longer need actual IR data.

A. Improved terrain classification

To improve the performance of terrain classification using RGB images, we integrate both RGB and virtual IR images. Figure 3 (c) shows the network architecture, which is also based on AdapNet++. The loss function of this network is defined as $\mathscr{L}_{CE}$ of Eq. 3.

B. Thermal inertia estimation

Thermal inertia, I, which is defined by the following, provides another aspect of the target terrain characteristics:

$$I = \sqrt{k_c \rho c_p}, \tag{4}$$

where k_c, ρ, and c_p are the thermal conductivity, the density, and the specific heat capacity of the soil, respectively.

To calculate the thermal inertia of the terrain, we need to solve the one-dimensional heat conduction equation:

$$\frac{\partial T(z,t)}{\partial t} = \frac{k_c}{\rho c_p} \frac{\partial^2 T(z,t)}{\partial z^2}, \tag{5}$$

where $T(z,t)$ is the temperature at time t and depth z, where $z = 0$ is at the terrain surface. Using the thermal inertia definition (Eq. 4), Eq. 5 can be rewritten as follows:

$$\frac{\partial T(z,t)}{\partial t} = \left(\frac{I}{\rho c_p}\right)^2 \frac{\partial^2 T(z,t)}{\partial z^2}. \tag{6}$$

To calculate the equation above, the heat input and output (heat budget) balances at the terrain surface, which is the upper boundary condition [20], as follows:

$$(1-A) Q_{SW} cos(\theta) + Q_{LW} - \varepsilon \sigma T_s^4 - Q_H - Q_E = -k_c \left. \frac{\partial T(z,t)}{\partial z} \right|_{z=0}, \tag{7}$$

where A is the albedo of the terrain; θ is an angle from surface normal to the Sun; Q_{SW} and Q_{LW} are the short-wave and the long-wave heat fluxes, respectively; ε is the

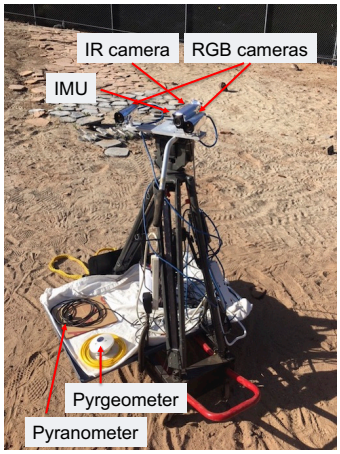

Fig. 4. System setting.

emissivity of the terrain, σ is the Stefan-Boltzmann constant (5.670×10^{-8} [$\mathrm{Wm^{-2}K^{-4}}$]); T_s is the temperature at the terrain surface, that is, $T_s = T(0,t)$; Q_H and Q_E are the sensible and latent heat fluxes, respectively. Although the Q_H and Q_E calculation would be beneficial for an Earth environment, we have ignored these towards simplification. For the lower boundary condition at $z = z'$, we set the constant value $T(z',t) = 200$ [K] which is a sufficiently lower temperature than the lowest temperature of the day.

To solve each parameter including the thermal inertia, the following optimization problem is solved:

$$\begin{aligned} \min_{A,\theta,\varepsilon,I,\rho c_p} \quad & \left[(T(0,t) - T_s)^2\right]. \\ \text{s.t.} \quad 0.0 \le A \le 1.0, \quad & 0.9 \le \varepsilon \le 1.0 \\ 50.0 \le I \le 800.0, \quad & 9.5e5 \le \rho c_p \le 1.2e6 \end{aligned} \tag{8}$$

where $T(0,t)$ is the *calculated* surface temperature, while T_s is the *measured* surface temperature by an IR camera. To calculate the model, we used the measured short-wave radiation Q_{SW} and long-wave radiation Q_{LW} from a pyranometer and pyrgeometer, respectively.

To estimate thermal inertia using the above equations, time-series data as shown in Fig. 2 is necessary. With an assumption of a constant albedo in each terrain type and size [21], we estimate thermal inertia at each terrain type by using time-series virtual IR images and estimated terrain types. More concretely, at each terrain type, the mean value of thermal, angle, pyranomter, and pyrgeometer information is calculated at each frame. The time-series thermal, angle, pyranomter, and pyrgeometer information at each terrain are used to estimate thermal inertia.

V. Experiments

In this section, we talk about the data collection and experimental results of the virtual sensor, the proposed terrain classification, and the thermal inertia estimation.

A. Data collection

We collected data at JPL Mars Yard with two RGB cameras (FLIR spinview), an IR camera (FLIR AX65), a pyranometer (Apogee/SP-420), a pyrgeometer (Kipp Zonen/SGR3), and an inertial measurement unit (IMU, SparkFun Electronics 9DoF Razor IMU) from 10:30 to 17:30 on Aug 2nd, 2019. The system with sensors is shown in Fig. 4. Data was collected every one hour by changing the position of the system in the Mars Yard, around 30 times every one hour (i.e. ∼30 sets of RGB and IR images and solar radiation collected once every hour).

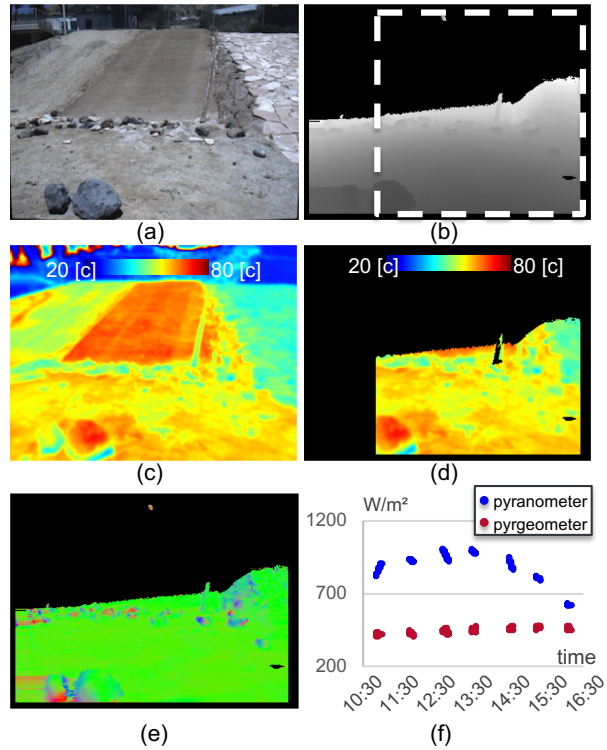

Fig. 5. (a) An example of a captured image, (b) depth image of (a), (c) IR image corresponding to (a), (d) projected IR image on (a), (e) angle image showing an angle from surface normal to the Sun at each pixel (each color channel is assigned to x, y, z values in angle image), and (f) time-series data of measured values by pyranometer (blue) and pyrgeometer (red).

Figures 5 (a) ∼ (c) show examples of a captured RGB image, its depth image, and IR image. We found that 3D points after 7 [m] and those of the left quarter image area are relatively noisy, so we decided to use 3D points within 7 [m] from the camera and the dotted rectangle area on Fig. 5 (b). The proposed neural network of the virtual sensor assumes that RGB and IR images are aligned. Therefore we first back-project an IR image on its 3D points, followed by projection on the RGB image. An example of aligned IR images is shown in Fig. 5 (d). Since alignment uses 3D points, IR information without 3D points are eliminated. Figure 5 (e) highlights an angle image showing an angle from the surface normal to the Sun at each pixel (each color channel is assigned to x,y,z values). Short wave and long wave radiation measured by the pyranometer and pyrgeometer, respectively, are shown in Fig. 5 (f).

For the purpose of terrain classification, we annotated RGB images. We labeled terrain into six classes numbered

TABLE I

PERFORMANCE COMPARISON BASED ON MAE (MEAN ABSOLUTE ERROR) OF VIRTUAL SENSOR WITH 4 INPUT SETTINGS ((1) RGB IMAGES, (2) RGB AND ANGLE IMAGES, (3) RGB IMAGES, AND (4) ALL).

Input	MAE [degree]
(1) RGB images	4.556
(2) RGB and angle images	4.367
(3) RGB images and solar radiations	3.394
(4) All	2.423

as follows: (1) rocks, (2) bedrock (red), (3) bedrock (black), (4) compacted sand with gravel, (5) compacted sand, and (6) loose sand with gravel. Figure 6 (a) shows the six terrain types in Mars Yard. Figures 6 (b) and (c) show examples of annotated images and those original RGB images. Each of (2) bedrock (red), (3) bedrock (black), (4) compacted sand with gravel, and (5) compacted sand are not widely distributed like (1) rocks and (6) loose sand with gravel in Mars Yard. This results in a very small number of data points at every data collection (e.g. only one or two data points for (2) bedrock (red), (3) bedrock (black), (4) compacted sand with gravel, and (5) compacted sand every 1 hour).

In the following experiments, we generated training, validation, and test data sets as follows. At each data collection time, 50%, 25%, and 25% of the data set are randomly selected. To train deep learning models, we used all of the training data set from 10:30 to 16:30. Validation and test are done using the validation and test data set from 10:30 to 16:30. Due to the limited amount of data, we note it is possible that a test sample collected at a certain location (e.g. just in front of bedrock (red)) at 10:30 might have an extremely similar sample included in the training set at a different time (14:30). In other words, if we collected data at 20 different locations in Mars Yard each hour, it is possible that a training sample taken at location 1 from the 10:30 collection has the similar geometric properties as a test sample at location 1 collected at 14:30.

B. Virtual IR sensor

In this section we evaluate the improved virtual IR sensor with the newly collected data. First we used the Multi-AdapNet++ with two outputs (IR and annotation images) as shown in Fig. 3 (b). We changed the input sensor as (1) RGB images, (2) RGB and angle images, (3) RGB images and solar radiations, and (4) all sensors. Figures 7 (a) $\sim$ (f) show visualization of an example of actual IR images taken at 15:30, its RGB image, and virtual IR and absolute error images with four input types, (c) RGB images, (d) RGB and angle images, (e) RGB images and solar radiations, and (f) all, respectively.

Table I shows the performance comparison of the four input settings. RGB images do not produce good results since it does not account for geological information, such as Sun angle, and radiation information. In general, we find that using more sensors leads to better performance.

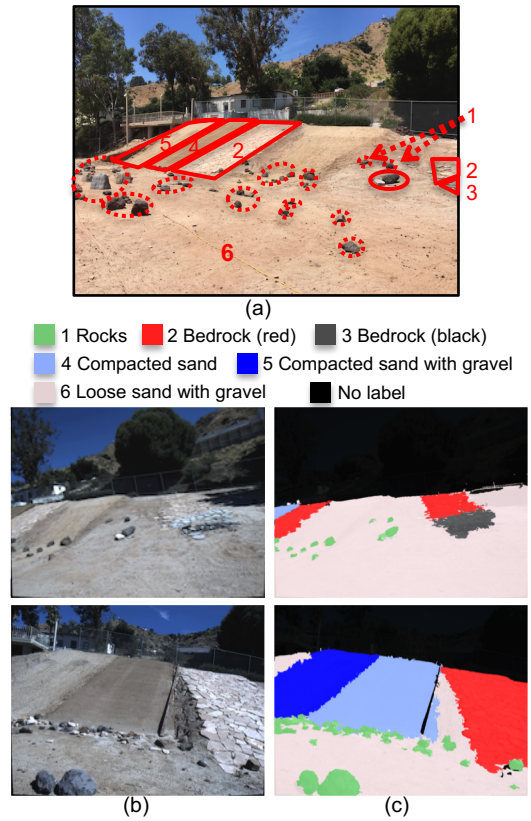

Fig. 6. (a) Six terrain types in Mars Yard ((1) rocks (red dotted circles), (2) bedrock (red), (3) bedrock (black), (4) compacted sand with gravel, (5) compacted sand, and (6) loose sand with gravel), (b) examples of captured images, and (c) annotated images corresponding to (b).

TABLE II

TERRAIN CLASSIFICATION RESULTS FROM (A) RGB IMAGES ONLY, (B) RGB AND ACTUAL IR IMAGES, (C) RGB AND VIRTUAL IR IMAGES BASED ON RGB AND SOLAR RADIATIONS, AND (D) RGB AND VIRTUAL IR IMAGES BASED ON ALL SENSORS.

Setting	Frequency weighted IoU	Mean IoU	Pixel accuracy
(a) RGB	0.893	0.483	0.935
(b) RGB and actual IR	0.915	0.522	0.953
(c) RGB and virtual IR based on RGB images	0.913	0.523	0.950
(d) RGB and virtual IR based on all sensors	0.914	0.522	0.951

To confirm the effectiveness of the use of two outputs (IR and annotation images) of the Multi-AdapNet++, we applied the Multi-AdapNet++ with single output (IR images) (Fig. 3 (a)). When using our best model, i.e. all sensors are used as input, the MAE (mean absolute error) is 5.02 [degree]. Compared with the result 2.423 [degree] in Table I (4), the single output model has much higher MAE, proving the effectiveness of the two-output architecture.

C. Terrain classification

We did four terrain classification experiments using (a) RGB images only, (b) RGB and actual IR images, (c)

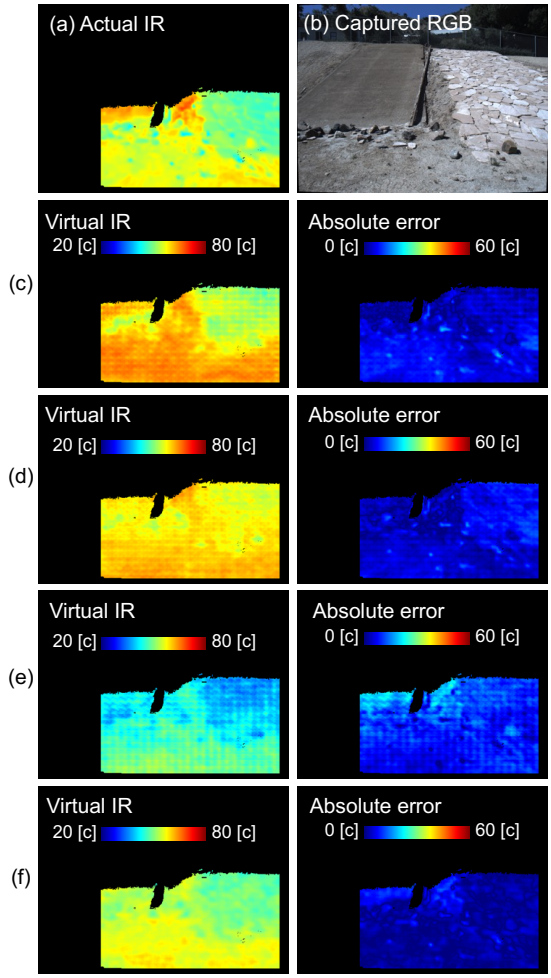

Fig. 7. (a) An example of actual IR images taken at 15:30, (b) RGB image of (a), (c)∼(f) virtual IR and absolute error images with input (c) RGB images, (d) RGB and angle images, (e) RGB images and solar radiations, and (f) all.

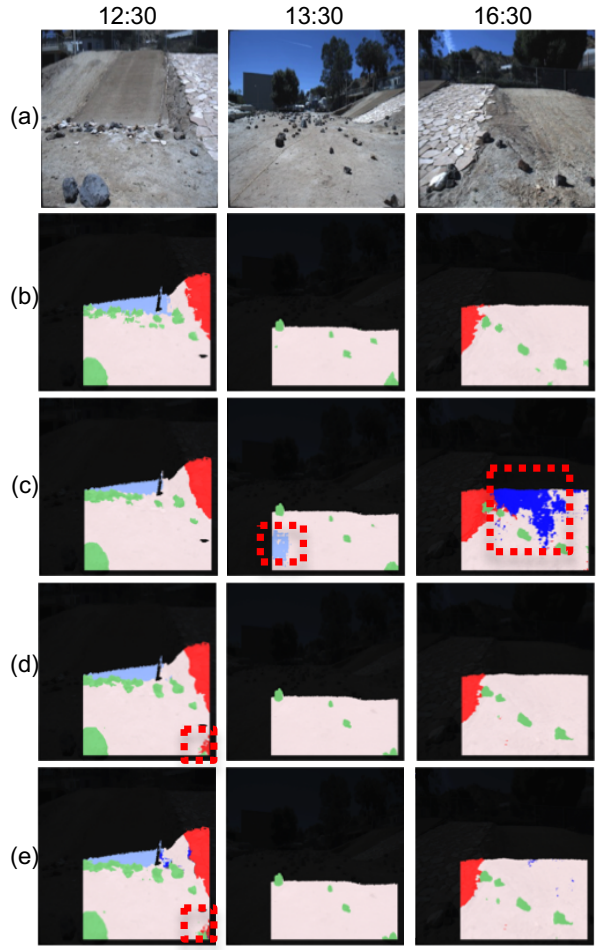

Fig. 8. Examples of (a) captured RGB images, (b) ground truth of terrain types, (c) terrain classification results using RGB images, (d) those using RGB and actual IR images, and (e) those of RGB and virtual IR images. Red dotted lines show false positives.

RGB and virtual IR images based on RGB images, and (d) RGB and virtual IR images based on all sensors. For (a) RGB images only, we applied AdapNet++ [19], and for (b), (c), and (d) we applied the proposed Multi-AdapNet++ (Fig. 3 (c)). Table II shows the results (frequency weighted intersection over union (IoU), mean IoU, and pixel accuracy) of the four experiments. Results in Table II (a) and (b) show the effectiveness of the use of IR information in addition to RGB information. The results of (c) and (d) show almost the same performance. One of possible reasons is that the absolute error at each pixel in virtual IR images does not have great impact on terrain classification. Relative temperature values between terrain types matter more. This is because the training data set consists of RGB and IR images, and it does not have any solar radiation information which is linked to terrain temperature. Therefore, future work includes training a model from RGB, IR images, and solar radiations. Lastly, the results of (b) are almost the same performance of (c) and (d). This shows the benefit of the virtual IR and highlights the possibility of increasing terrain classification performance of future rovers only equipped with only RGB cameras.

Although above results do not show the advantage of the use of all sensors in virtual sensor, we show that the use of all sensors is more robust to noise than the use of RGB images only in section V.D (sensitivity analysis).

Figure 8 shows examples of captured RGB images, ground truth terrain types, terrain classification results using RGB images, those using RGB and actual IR images, and those of RGB and virtual IR images based on all sensors, respectively, from top to bottom. Red dotted lines show false positives, and these suggest that the use of IR images in addition to RGB images is able to make the terrain classifier robust.

Table III shows a confusion matrix of terrain classification results from RGB and virtual IR images. The classes of ”bedrock (black)” and ”compacted sand with gravel” have a very limited number of pixels, due to a sensor location issue during data collection. The class of ”bedrock (black)” is mis-classified as rocks and bedrock (red), due to the very small number of training data and similarity with other rocks. The class of ”compacted sand with gravel” has only a single image at each data collection, and this results in zero images in the test data.

In the context of Moon or Mars the dangers of the

TABLE III

CONFUSION MATRIX OF TERRAIN CLASSIFICATION RESULTS FROM RGB AND VIRTUAL IR IMAGES. CLASSES (1) ∼ (6) SHOW (1) ROCKS, (2) BEDROCK (RED), (3) BEDROCK (BLACK), (4) COMPACTED SAND WITH GRAVEL, (5) COMPACTED SAND, AND (6) LOOSE SAND WITH GRAVEL.

Classes	Rocks	Bedrock (red)	Bedrock (black)	Compacted sand with gravel	Compacted sand	Loose sand with gravel
Rocks	0.883	0.001	0.0	0.0	0.002	0.108
Bedrock (red)	0.008	0.947	0.002	0.002	0.0	0.041
Bedrock (black)	0.509	0.311	0.029	0.0	0.0	0.149
Compacted sand with gravel	N/A	N/A	N/A	N/A	N/A	N/A
Compacted sand	0.039	0.0	0.003	0.025	0.833	0.100
Loose sand with gravel	0.015	0.009	0.0	0.001	0.001	0.971

terrain often come from obstacles and slopes that need to be avoided, rocky regions which potentially can damage the wheels (which has happened, as rocks perforated one of the wheels of the Curiosity rover), sandy regions, which can create slip, sinking, and even restrict ability to move further, also in case of the Spirit rover. Therefore, higher accuracy of terrain classification is desirable. The improvement which we showed in this section is important to success of the mission.

D. Sensitivity analysis

In this section we evaluate the proposed virtual sensor and terrain classification method in terms of robustness to noise. We used input sensors to the virtual sensor as (1) RGB only and (2) all sensors. We added Gaussian noise RGB and angle images for (2) as follows. The mean and standard deviation for Gaussian noise are defined as 0 and X, where X is percentage of standard deviation of all pixel values in RGB images in training dataset, respectively. Here, we used X=0.1, 0.25, 0.5, 0.75, and 1.0. We added Gaussian noise to angle images in the same manner with RGB images. Tables IV (a) and (b) show the mean absolute errors (MAE) of (1) RGB only and (2) all sensors by changing X. These results show that more sensors produce better results. Examples of RGB images after adding noise and estimated virtual IR images usgin all sensors are shown in Figs. 9 (a) and (b). The original image (X=0) and actual IR image are shown in Figs. 5 (a) and (d).

Next, the virtual IR images and RGB images, to which we add noise in the same way as the virtual sensor, are used to classify terrain types. Results are shown in Table IV (c) and (d). These results also show better performance of all sensors than RGB only. When X is more than 0.5, classification becomes harder even for humans. The accuracy of X=1.0 is more than 77%, which is in fact above our expectation when humans assess the images visually (based on authors' own assessments). Examples visualizations of classified terrain types using all sensors are shown in Fig. 9 (c).

E. Thermal inertia estimation

In general 24 hours of data is used for thermal inertia estimation, although we have only six hours of data (from 10:30 to 16:30). To show the reliability of the use of six hours data, first we use 24 hours data. We collected 24 hours data at JPL, and the target terrain is sand as shown in Fig. 10. Here, we used actual IR images for thermal inertia estimation. Estimated thermal inertia values using 24 hours data and six hours data (from 10:30 to 16:30) are 258 and 236, respectively, and the error is ∼8 %.

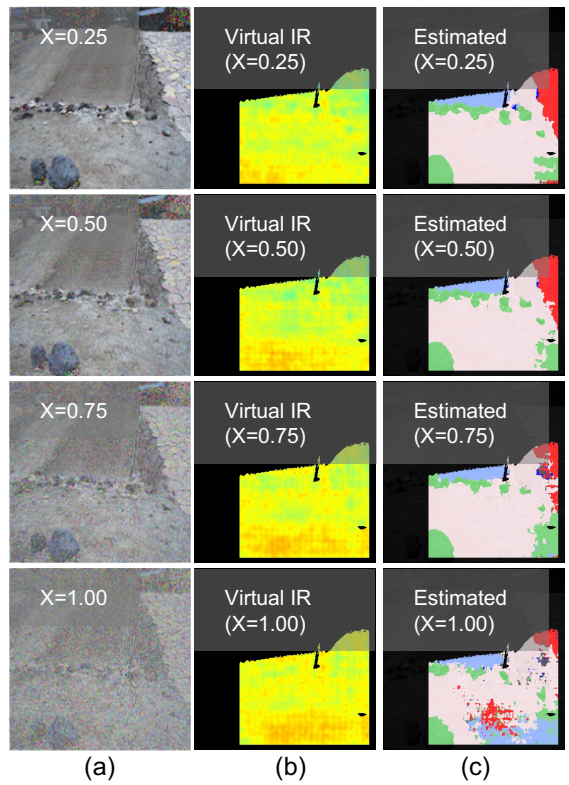

Fig. 9. (a) Examples of RGB images after adding Gaussian noise (X=0.25, 0.5, 0.75, and 1.0), for which mean and standard deviation are 0 and X, where X is percentage of standard deviation of all pixel values in RGB images in training data, (b) virtual IR images using all sensors, and (c) estimated terrain types using all sensors.

Finally we used virtual IR images to estimate thermal inertia at each terrain. Table V (a) shows estimated values using virtual IR images and estimated terrain types. We also used actual IR images and ground truth terrain types as shown in Table V (b). Here, we show the results of rocks, bedrock (red), and loose sand with gravel since we could not use three terrain types (bedrock (black), compacted sand, and compacted sand with gravel) due to very limited number of images at every one hour data collection. Although errors

TABLE IV

(A) AND (B) MAEs (MEAN ABSOLUTE ERRORS) OF VIRTUAL SENSOR BY RGB ONLY AND ALL SENSORS, RESPECTIVELY, AND (C) AND (D) PA (PIXEL ACCURACY) OF TERRAIN CLASSIFICATION RESULTS OF RGB AND VIRTUAL IR BY RGB ONLY AND ALL SENSORS, RESPECTIVELY.

X	(a) MAE (RGB only)	(b) MAE (all sensors)	(c) PA (RGB only)	(d) PA (all sensors)
0.1	4.597	2.393	0.946	0.945
0.25	5.438	2.580	0.929	0.941
0.5	5.594	2.873	0.908	0.909
0.75	5.152	2.903	0.838	0.857
1.0	4.973	3.028	0.772	0.792

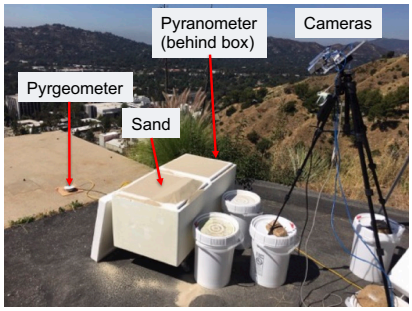

Fig. 10. Experimental setting for 24 hours data collection.

are relatively high (varies from 14 % to 24 %), these results show a possibility of the use of virtual IR sensor for thermal inertia estimation. Moreover, the bedrock value shows a much smaller value compared with the bedrock in Fig. 2 (thermal inertia is 2506 with 20 cm skin depth). This suggests that there are loose terrain area (such as sand) under the bedrock, which is the case in the Mars Yard.

VI. CONCLUSIONS

We proposed an improved virtual sensor by utilizing the angle from the surface normal to the Sun and the measurement of solar radiation to reduce the effect of slope angle variations and Sun angle. We proposed Multi-AdapNet++ to fuse multiple sensor inputs (RGB images, angle and solar radiation information) efficiently. We applied the virtual IR sensor to two applications: (1) improved terrain classification and (2) thermal inertia estimation. Using a newly collected data set, we verified the improved virtual IR sensor and both applications.

Results in terrain classification experiments show that absolute error in virtual IR images do not have much affect on terrain classification performance. Future work includes more detailed evaluations, such as training a model using RGB, IR image, and solar radiations.

The data set we collected was taken for only one day. We will extend our current data set into multiple days, multiple seasons, and multiple places to generalize the data set, followed by evaluation of proposed approaches with the new data sets.

TABLE V

THERMAL INERTIA ESTIMATION USING (A) VIRTUAL IR IMAGES AND (B) ACTUAL IR IMAGES AT EACH TERRAIN TYPE (ROCKS, BEDROCK (RED), COMPACTED SAND, LOOSE SAND WITH GRAVEL.

	(a) Virtual IR images	(b) Actual IR images
Rocks	236	280
Bedrock (red)	449	581
Loose sand with gravel	282	246

ACKNOWLEDGEMENT

The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES

[1] B. Rothrock and et al., "Spoc: Deep learning-based terrain classification for mars rover missions," *AIAA SPACE*, pp. 5539—-5551, 2016.

[2] Y. Iwashita, K. Nakashima, A. Stoica, and R. Kurazume, "Tu-net and tdeeplab: Deep learning-based terrain classification robust to illumination changes, combining visible and thermal imagery," *IEEE Conf. on Multimedia Information Processing and Retrieval*, 2019.

[3] Y. Iwashita, K. Nakashima, S. Rafol, A. Stoica, and R. Kurazume, "Mu-net: Deep learning-based thermal ir image estimation from rgb image," *IEEE Workshop on Perception Beyond the Visible Spectrum (PBVS)*, 2019.

[4] C. Cunningham, I. Nesnas, and W. Whittaker, "Improving slip prediction on mars using thermal inertia measurements," *Autonomous Robots*, vol. 43, no. 2, 2019.

[5] S. Golombek, M.and Piqueux and et al., "Geology and physical properties investigations by the insight lander," *Space Sci. Rev.*, 2018.

[6] S. Gupta, J. Hoffman, and J. Malik, "Cross modal distillation for supervision transfer," *CVPR*, 2016.

[7] N. Srivastava and R. R. Salakhutdinov, "Multimodal learning with deep boltzmann machines," *NIPS*, 2012.

[8] D. Xu, W. Ouyang, E. Ricci, X. Wang, and N. Sebe, "Learning cross-modal deep representations for robust pedestrian detection," *CVPR*, 2017.

[9] L. Zhang and et al., "Synthetic data generation for end-to-end thermal infrared tracking," *IEEE Trans. on Image Processing*, no. 4.

[10] K. Yun and et al., "Improved visible to ir image transformation using synthetic data augmentation with cycle-consistent adversarial networks," *arXiv:1904.11620*, 2019.

[11] S. Salamati and et al., "Semantic image segmentation using visible and near-infrared channels," *ECCV 2012. Workshops and Demonstrations*, 2012.

[12] P. Filitchkin and et al., "Feature-based terrain classification for littledog," *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2012.

[13] I. Halatci and et al., "Terrain classification and classifier fusion for planetary exploration rovers," *IEEE Aerospace Conference*, 2007.

[14] L. Chen and et al., "Semantic image segmentation with deep convolutional nets and fully connected crfs," *PAMI*, 2017.

[15] A. Valada and et al., "Robust semantic segmentation using deep fusion," *Robotics: Science and Systems (RSS 2016) Workshop Are the Sceptics Right? Limits and Potentials of Deep Learning in Robotics*, 2016.

[16] S. Chhaniyara and et al., "Terrain trafficability analysis and soil mechanical property identification for planetary rovers: A survey." *J. of Terramechanics*, vol. 49, no. 2, p. 115–128, 2012.

[17] N. E. Putzig, "Thermal inertia and surface heterogeneity on mars," *PhD thesis, University of Colorado*, 2006.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *MICCAI*, 2015.

[19] A. Valada, R. Mohan, and W. Burgard, "Self-supervised model adaptation for multimodal semantic segmentation," *Int. J. of Computer Vision (IJCV)*, 2019, special Issue: Deep Learning for Robotic Vision.

[20] G. M. Martínez and et al., "Surface energy budget and thermal inertia at Gale Crater: Calculations from ground-based measurements," *J. of Geophysical Research: Planets*, vol. 119, no. 8, pp. 1822–1838, 2014.

[21] S. Nowicki and et al., "Rock abundance on mars from the thermal emission spectrometer," *Geophysical research planets*, 2007.