Prediction of Vibrations as a Measure of Terrain Traversability in Outdoor Structured and Natural Environments

Mohammed Abdessamad Bekhti$^{(\boxtimes)}$ and Yuichi Kobayashi

Graduate School of Science and Technology, Shizuoka University, Hamamatsu, Japan
bekhti.m.a@gmail.com, kobayashi.yuichi@shizuoka.ac.jp

Abstract. Terrain recognition is an important task that a mobile robot has to accomplish autonomously to navigate in hazardous territories safely with no additional human monitoring. For this, sensory information should be employed to construct a good model to estimate the degree of traversability of upcoming terrains. In this paper, a regression-based method is proposed to estimate mobile robot vibration from terrain images as a description for terrain traversability. Texture attributes obtained from evaluation of the fractal dimension to describe the terrains were combined with appropriate acceleration features for function approximation using Gaussian Process regression (GP). Results showed effectiveness of the method to predict motion data for different terrain configurations in structured and rough environments.

Keywords: Autonomous mobile robots · Texture features · Acceleration · Gaussian process regression · Terrain traversability

1 Introduction

Autonomous mobile robot navigation in natural environments is a challenging task, due to variety of environment configurations a mobile platform may encounter and lack of a precise prior knowledge of the terrain conditions [1]. Thereby, it is important to design a highly autonomous recognition system that will allow the mobile platform to rely less on human instructions to assess terrain traversability, *i.e.*, terrain representation can evolve through interaction with new terrain scenarios without referring to predefined set of terrain classes such as grass, rocks, mud, etc.

Several sensors to perceive the environment have been used to develop frameworks for traversability. They can be categorized as exteroceptive and proprioceptive sensors [2]. Exteroceptive sensors such as cameras and lasers give a direct representation of the environment. In addition to this, wheel encoders and proprioceptive sensors such as vibration sensors, current sensors, gyroscopes provide complementary information about the environment, and thus the importance of combining both type of sensors to improve terrain recognition and reducing the dependency on human design.

A speed control system was developed based on roughness identification using a 2D laser [3]. Terrain scans return the height of the point on the soil which are compared with measures of flatness obtained by scanning the height of a flat surface in front of vehicle. Based on the comparison of the four parameters used to describe the terrain roughness with predefined thresholds, the speed is controlled to be either reduced, augmented or maintained at the current level.

Shape attributes of the ground were extensively investigated for traversability estimation. The idea is to use 3D environment analysis methods and chart the mobility smoothness in a hierarchical fashion. Based on this, 3D Ladar information was clustered according to scatterness (grass), surfaceness (ground), and linearness (tree trunc) for terrain modeling [4]. 2D lasers are plane scanners and thus unable to describe the geometry of the soil, also obstacle detection (e.g., [5]) may fail if the scanned plane is high. On the other hand, 3D lasers or lidars provide detailed representation of the scenery [6], but still are costly, consume more energy to perform.

Cameras are undoubtedly more preferred since they are much affordable, easier to integrate, and can furnish various information such as color, texture, etc. In this context, several methods have been implemented methods which used both image and geometry data for traversability estimation.

Investigating elevation of height profile allowed detection and categorization of obstacles into negative (ditches, holes) and positive (bump) obstacles [7]. Terrain categorization was done through classification of color information within predefined material-class. Obstacle degree of traversability was inferred by combining range information and classification outcome. In the same context, an algorithm for self-supervise terrain classification combining geometrical and visual features was proposed [8]. With regard to image information, HSV color expressed by a four dimensional vector and wavelet-based texture information were used. Features are then fused using a naive Bayes classifier. Both previously mentioned methods lack use of low level feedback information (proprioceptive data) related to the moving platform that can be used as ground truth.

In the context of proprioceptive data, vibrations generated from interaction of the wheels and the soil was used to perform terrain classification [9]. The acceleration signal is divided into small sections which will serve to evaluate the Power Spectral Density (PSD). To categorize terrain samples Linear Discriminant Analysis (LDA) was employed. Vibration data may provide a good understanding of robot behavior while navigating but it is not enough to anticipate upcoming terrains and scenarios such as obstacles since it is limited to underfoot only.

Latest research projects showed the effectiveness of learning from both exteroceptive and proprioceptive sensory information to estimate the vehicle behavior toward subsequent terrains. Image color information was employed by Krebs *et al.* to approximate the interaction of the mobile platform and the ground based on a classification framework [10]. The goal of training process was to learn terrain understanding through interaction with several terrain configurations. Rover Terrain Interaction (RTI) attributes were combined with corresponding

exteroceptive data to build an inference model. It showed effectiveness for RTI parameters such as slip and vibration.

Regression of terrain traversability based on robot pose was proposed by Ho *et al.* [11]. Vehicle attitude was estimated when running on deformable terrains from an approximation done on rigid grounds and terrain geometry attributes using Gaussian Process regression (GP) [15]. For the final objective of navigation, geometry is just an intermediate representation. It is possible to directly predict motion property from other sensory modalities like cameras which can provide 3D map of the environment.

Image texture information can furnish a detailed representation about terrain natures due to the high dimension of the feature vector compared with color information [10], and thus enables a good prediction for traversability. Nonetheless, classifying terrain using only texture attribute may not be suitable. Another issue in classification framework is the non-consistency of attributed labels, since it relies on the judgment of the designer.

This paper addresses approximation of terrain signature *i.e.*, vibrations from mobile robot motion information based on learned data association of exteroceptive and proprioceptive information using GP regression. With regard to exteroceptive data, texture information based on fractal dimension is used to describe terrain configurations. Also vibration *i.e.*, vertical acceleration generated from interaction between the mobile platform and the soil was considered in this study case. Compared with semi-supervised methods, the proposed framework uses acceleration signals obtained from interaction of the vehicle with the soil as ground truth. The proposed approach is able to estimate the vibration that the mobile platform may encounter when running on a terrain regardless of its nature.

2   Problem Definition

2.1   Objective

A camera and an acceleration sensor are mounted on a four wheel mobile platform as shown in Fig. 1. The camera acquires terrain images and the acceleration sensor registers acceleration signal generated during a sequence of a run. The target is to predict vibration feature as a measure of traversability only from terrain images. This information will be used to understand the vehicle behavior towards subsequent terrains and is expected to enable a safe run.

2.2   Methodology

From terrain images, Region of Interest (ROI) are determined for image feature extraction. In this case study, from every single image two terrain patches (ROIs) will be used. Texture attribute evaluated by measuring the fractal dimension [13] is retained. Registered acceleration signal is divided into small sequences to measure motion features according to extracted terrain patches. For function
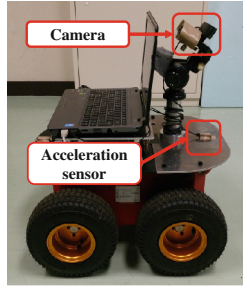
Fig. 1. Pioneer 3AT hardware configuration.

approximation, GP regression is employed to learn data correlation of texture and vibration features. Based on this result, prediction of vibration features is performed using only image information.

3 Regression of Motion Information

3.1 Terrain Patches Identification

From terrain images, patches corresponding to a short range distance will be considered for texture extraction. For this, the camera model is used to define the ROIs (refer to [12] for further details). As shown in Fig. 2, the camera is characterized by the camera frame $\{O_c, x_c, y_c, z_c\}$ whose center is O_c. The following equation translates a point from the world reference frame represented by $P \in \mathbb{R}^4$ to $p \in \mathbb{R}^3$ (both containing one as the last element) onto the image plane:

$$p = \mathbf{K}[\mathbf{R}|t]P, \tag{1}$$

where $[\mathbf{R}|t] \in \mathbb{R}^{4\times4}$ is the extrinsic parameter matrix, and $\mathbf{K} \in \mathbb{R}^{4\times3}$ is the camera parameter matrix.

The above-mentioned model will be used to determine terrain the vehicle is traversing using robot's wheels base L and width W. As shown in Fig. 3, both camera and world references frame are center of the platform. The world
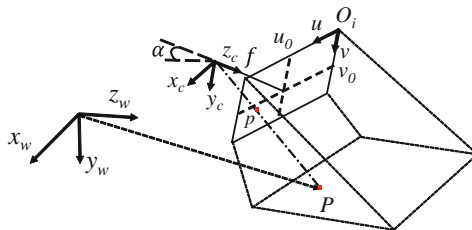

Fig. 2. Camera frame for mapping 3D point mapping onto the image plane.

reference frame is given by $\{O_w, x_w, y_w, z_w\}$ with O_w as origin. The goal of the process is to map points to identify the inner and outer bounds of terrain region crossed by the vehicle. Since all points lay on the same line, only the depth components according to the w_z-axis will vary. For both left and right sides, these points are expressed as follows:

$$
\begin{cases}
P_{ro} = [\frac{L}{2} + W, & 0, & z_w] \\
P_{ri} = [\frac{L}{2}, & 0, & z_w] \\
P_{lo} = [-\frac{L}{2} - W, & 0, & z_w] \\
P_{li} = [-\frac{L}{2}, & 0, & z_w]
\end{cases}
, \tag{2}
$$

where P_{ro}, P_{ri} denote the points laying on the right outer and inner bounds respectively and P_{lo}, P_{li} the left outer and inner bounds. Results of this operation are shown in Fig. 4. To increase the number of terrain sample, two terrain patches will be extracted from a single image. Warp perspective transformation is performed to remove the trapezoidal shape introduced when taking a picture.

3.2   Texture Attribute Extraction

Texture attributes are extracted using Segmentation-based Fractal Texture Analysis (SFTA) [13]. The SFTA method consists of two steps. In the first step, the input grayscale image is decomposed into binary images using the Two-Threshold Binary Decomposition (TTBD) [13]. In the second step, each binary image is used to calculate the fractal dimension from its region boundaries. TTBD computes a set of thresholds T, which is acheived by the Multi-level Otsu algorithm [14]. After obtaining the desired threshold number n_t, pairs of contiguous thresholds $T \cup \{n_l\}$ with n_l the highest value in the input gray scale image, associated with pairs of thresholds $\{t, n_t\}$ with $t \in T$, are employed to obtain the binary images as follows

$$
I^b_{x,y} = \begin{cases} 1, & \text{if } t_l < I(x,y) \le t_u \\ 0, & \text{otherwise} \end{cases}, \tag{3}
$$

where $I^b_{x,y}$ denotes the binary value of image pixel (x, y), t_l and t_u denote the lower and the upper threshold values respectively. Consequently, $2n_t$ binary
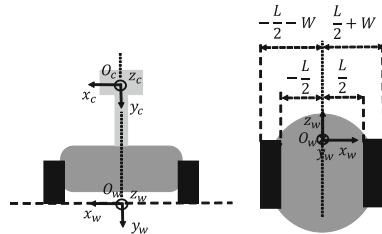
Fig. 3. Mobile robot carrying a camera and layout of both world and camera reference frames.

Fig. 4. Identification of terrain patches for image feature extraction.

images will be obtained. In the application of this paper, SFTA feature vector will include only the fractal dimension of the boundaries. $\Delta(x, y)$ represents the border image of the binary image $I^b_{x,y}$, and is evaluated as

$$\Delta(x, y) = \begin{cases} 1, & \text{if } \exists (x', y') \in N_{x,y} \\ & s.t.\ I^b_{x',y'} = 0 \wedge I^b_{x,y} = 1 \ , \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $N_{x,y}$ is the 8-connexity of a pixel (x, y). $\Delta(x, y)$ takes the value one if the pixel at location (x, y) in the related binary image $I^b_{x,y}$ has the value one and having at least one neighbor pixel with zero value. The border image is used to compute the fractal dimension $D \in \mathbb{R}$ by the box counting algorithm as follows

$$D_0(x, y) = \lim_{\varepsilon \to 0} \frac{\log N(\varepsilon)}{\log(\varepsilon^{-1})}, \tag{5}$$

where $N(\varepsilon)$ is the number of hyper-cubes of length ε that fill the object. The resulting SFTA feature vector is denoted by $\boldsymbol{x} \in \mathbb{R}^{2n_t}$.

3.3   Motion Feature Extraction

Let $a_k, k = 1, \cdots, K$ be the vertical acceleration signal translating vibrations during a run on a short range distance with time step k and K is the total time steps. Amplitude distance was retained to describe the motion of the platform, and is given by

$$a_{k_d} = \max_{k=1,\cdots,K} a_k - \min_{k=1,\cdots,K} a_k. \tag{6}$$

3.4   Matching Image and Motion Features

As mentioned above, two short terrain segments are considered for image feature extraction. Accordingly, relevant acceleration segment for motion feature
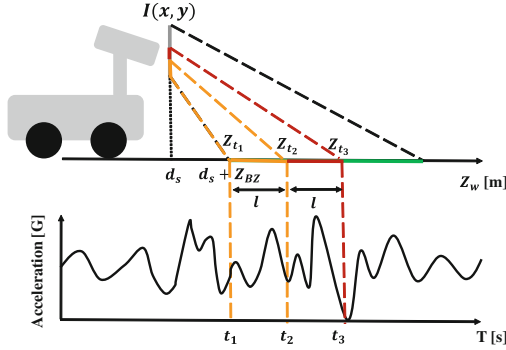
Fig. 5. data acquisition

extraction is matched to image features using coordinate transformation and odometry. As shown in Fig. 5, at time $t = 0$, the robot is located at the origin of the world reference frame. Location to take a new picture is denoted as $Z_{I_i(x,y)}$ and is expressed as

$$Z_{I_i(x,y)} = id_s, \quad i = 1, \cdots, N_{\text{image}}, \tag{7}$$

where d_s denotes sampling distance, and N_{image} is the total number of terrain images acquired during a run. Due to the camera tilt angle α, terrains will be covered from a certain position expressed as

$$Z_{I_i(x,y)} + Z_{\text{BZ}}, \quad i = 1, \cdots, N_{\text{image}}, \tag{8}$$

where Z_{BZ} is area not covered by the image, which is called blind zone. As mentioned above, a short range distance l covered by the images will be retained for image feature extraction. A reason for that is the further a pixel (x, y) is from the camera focus point, the more noise the pixel will be subject to. Thus visual representation of the environment may not be faithful to the real world. Acceleration signal sequence generated by the distance l is used for motion feature extraction, and is bounded according to the ROI as follows

$$A_i = [Z_{I_i(x,y)} + Z_{\text{BZ}}, Z_{I_i(x,y)} + Z_{\text{BZ}} + l], \tag{9}$$

where A_i denotes the ROI.

3.5   Regression Analysis using Gaussian Process (GP)

Gaussian process [15] is used for regression analysis. Let $\mathcal{D} = \{(\boldsymbol{x}_j, y_j), j = 1, \cdots, n\}$ denote a set of training samples with pairs of SFTA feature vector inputs $\boldsymbol{x}_j$ and motion feature outputs $y_j \in \mathbb{R}$, n denotes the total number of training samples. The goal is to evaluate the predictive distribution for the function f_* for test input samples $\boldsymbol{x}_*$. The noise is additive, independent and

(a) Gravel

(b) Stones

(c) Grass

(d) Woodchip
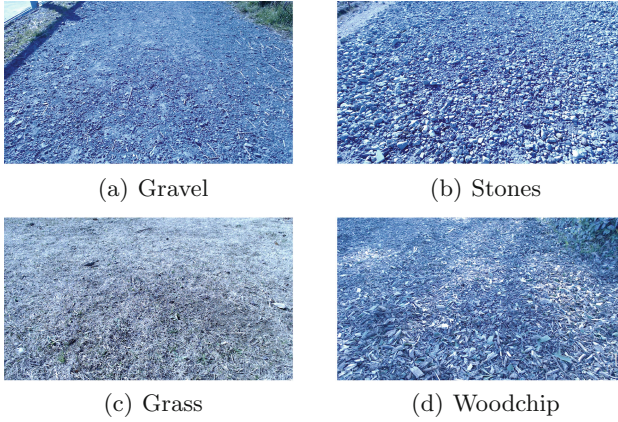
Fig. 6. Natural terrains.

following a normal distribution. The relationship between the function $f(\boldsymbol{x})$ and the observed noisy targets y is given by

$$y_j = f(\boldsymbol{x}_j) + \varepsilon_j, \tag{10}$$

where ε_j is a noise which follows $\mathcal{N}(0, \sigma^2_{\text{noise}})$, where σ^2_{noise} denotes the variance of the noise. The notation $\mathcal{N}(\mathbf{a}, A)$ is retained for the normal distribution with mean $\mathbf{a}$ and covariance A. Gaussian Process regression is a Bayesian algorithm that assumes that *a priori* the function values respond as

$$p(\mathbf{f}|\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n) = \mathcal{N}(0, K), \tag{11}$$

where $\mathbf{f} = [f_1, f_2, \cdots, f_n]^\top$ contains the latent function values, $f_j = f(\boldsymbol{x}_j)$ and K is a covariance matrix which inputs are provided by the covariance function defined as $K_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$. A widely employed covariance function is the squared exponential given by

$$K_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sigma^2 \exp\left(-\frac{(\boldsymbol{x}_i - \boldsymbol{x}_j)^\top (\boldsymbol{x}_i - \boldsymbol{x}_j)}{2\lambda^2}\right), \tag{12}$$

where σ^2 controls the variance, and λ is the isotropic lengthscale parameter that describes the smoothness of a function.

4   Experiment

4.1   Experiment Settings

To predict motion information based on Gaussian Process regression, the vehicle was engaged extensively in 30 different outdoor environment configurations

(a) Slik asphalt   (b) Granulated asphalt

(c) Tiles   (d) Wood
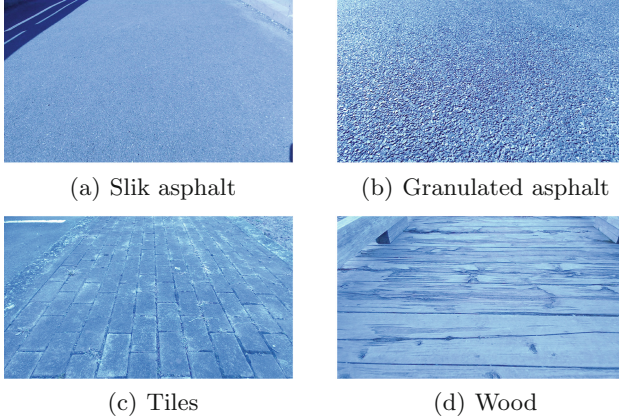
Fig. 7. Artificial terrains.

to acquire terrain observations (images) and corresponding acceleration signals. The goal of experiment is to assess performance of proposed framework to predict motion features. Samples of terrain configurations are shown in Figs. 6 and 7. Training was done using a set of 1208 terrain patches and corresponding acceleration features. An extra set of 109 image features were used for testing. Experiment parameters were set as follows

- Camera height h: 540 [mm],
- Camera tilt angle α: 31[deg],
- Number of thresholds for SFTA n_t: 8,
- Dimension of SFTA feature vector $\boldsymbol{x} \in \mathbb{R}^{16}$,
- ROI length l: 250 [mm],
- Number of images used for function approximation N_{tr}: 1208,
- Number of images used for prediction N_{pr}: 109,
- Acceleration signal sampling period T_s: 10 [ms],
- Step distance for taking images d_s: 500 [mm].

4.2   Results

Effectiveness of the proposed approach was evaluated based on whether the real motion feature belongs to 1σ trust interval. The results are introduced by Fig. 8. A total of 76 terrain samples of different natures were successfully approximated within the trust interval. Moreover, the fractal dimension offers a discrimination of terrains texture attributes. Principal Component Analysis (PCA) was applied to SFTA feature vector to reduce the dimension from $\boldsymbol{x} \in \mathbb{R}^{16}$ to $\boldsymbol{z} \in \mathbb{R}^2$ to visualize the distribution of the motion information. As shown in Fig. 9, terrain configurations are scattered in different areas.
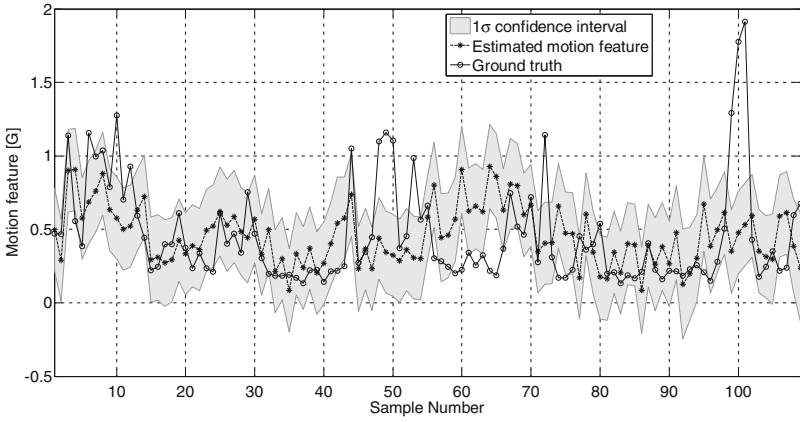
Fig. 8. Approximation of motion features using Gaussian process
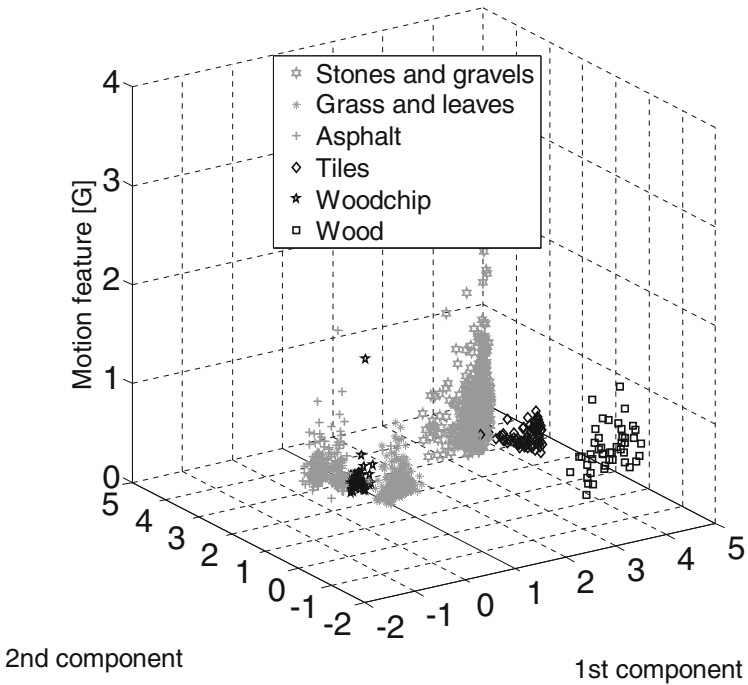

Fig. 9. Distribution of motion information with regard to SFTA for three terrain types: stones and gravels, grass and leaves, granulated and slick asphalt.
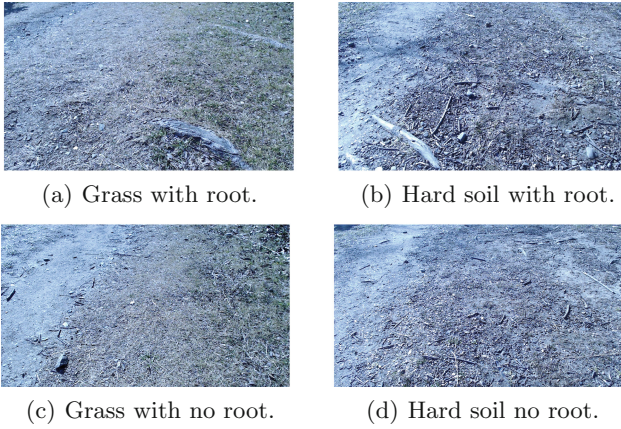
(a) Grass with root.

(b) Hard soil with root.

(c) Grass with no root.

(d) Hard soil no root.

Fig. 10. Difference of terrain configurations for same ground nature.
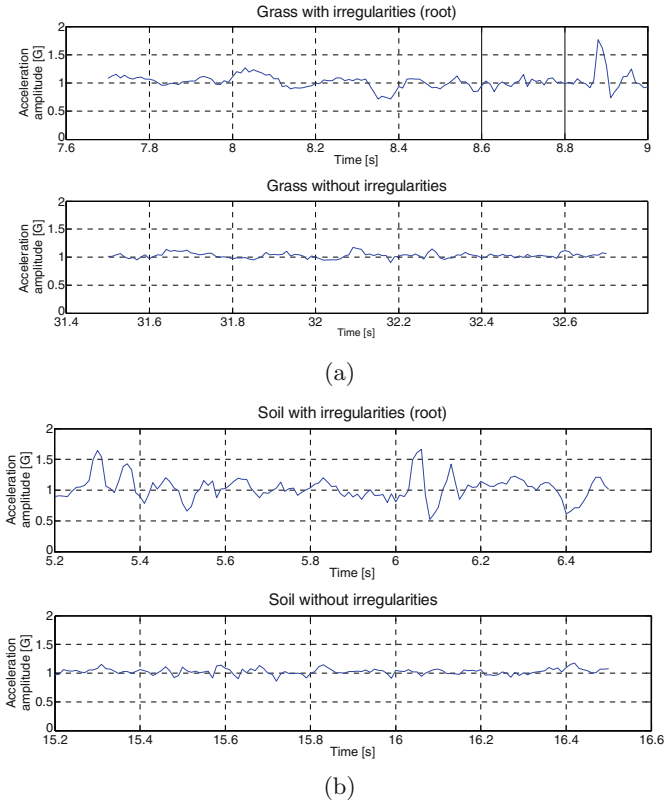

(a)

(b)

Fig. 11. Difference in vibration signal for terrains with and without irregularities: (a) Grass, (b) soil.

4.3   Discussion

The remaining 33 samples were not estimated within the confidence interval. In the case of terrain configurations shown by Fig. 10(a) and (b), where tree roots appear from the ground creating bumps, motion data could not be predicted well. When the mobile platform crosses terrain deformation or irregularity (tree roots, rocks, etc.), acceleration sensor acquires instant high vibration values *i.e.*, motion information is more localized compared with uniform terrains which generates smoother acceleration signal as shown in Fig. 11. On the other hand, image region can include both grass textures and bumps. More localized (with narrower regions) features to take into account irregularities of the soil will improve this problem. Thus, sufficient number of data will be required also for terrain irregularities to achieve a good approximation of motion features.

5   Conclusion and Future Work

In this paper, we introduced a method to assess terrain mobility for mobile robot navigating in outdoor environments based on data association of exteroceptive data (terrain observations) and proprioceptive information (vibrations of the platform). SFTA texture attribute was used to describe the nature of terrains. It can provide a discrimination of terrain texture information and thus the ability to distinguish terrain natures (stones, grass, and asphalt). Using GP regression, a set of training data of different terrain configurations was used for learning by combining both image and motion attributes, and a different set of terrain images was used for validation. Obtained results showed a good motion data prediction with 70 % success rate. SFTA feature vector is a non localized representation of terrain characteristics with regard to the running information which is more local as of its instantaneous nature. A more localized image representation may be suitable for predicting local motion data of the mobile robot when running on a terrain with irregularities. More general problems could influence the image features such as high brightness which may lead to a strong reflectance from the ground. Also, perspective effect of the camera leads to a less appreciation of the details included in the soil.

References

1. Chilian, A., Hirschmuller, H.: Stereo camera based navigation of mobile robots on rough terrain. In: IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4571–4576. IEEE Press (2009)
2. Sancho-Pradel, D., Gao, Y.: A survey on terrain assessment techniques for autonomous operation of planetary robots. J. Br. Interplanetary Soc. **63**, 206–217 (2010)

3. Castelnovi, M., Arkin, R., Collins, T.: Reactive speed control system based on terrain roughness detection. In: IEEE International Conference on Robotics and Automation, pp. 891–896. IEEE Press (2005)
4. Lalonde, J.-F., Vandapel, N., Huber, D.F., Hebert, M.: Natural terrain classification using three-dimensional ladar data for ground robot mobility. J. Field Robot. **23**, 839–861 (2006)
5. Labayrade, R., Gruyer, D., Royere, C., Perrollaz, M., Aubert, D.: Obstacle Detection Based on Fusion Between Stereovision and 2D Laser Scanner. Mobile Robots: Perception and Navigation (2007)
6. Papadakis, P.: Terrain traversability analysis methods for unmanned ground vehicles: a survey. Eng. Appl. Artif. Intel. **26**, 1373–1385 (2013)
7. Bellutta, P., Manduchi, R., Matthies, L., Owens, K., Rankin, A.: Terrain perception for DEMO III. In: IEEE Intelligent Vehicles, pp. 326–331. IEEE Press (2000)
8. Brooks, C.A., Iagnemma, K.D.: Self-supervised classification for planetary rover terrain sensing. In: IEEE Aerospace Conference, pp. 1–9. IEEE Press (2007)
9. Brooks, C., Iagnemma, K., Dubowsky, S.: Vibration-based terrain analysis for mobile robots. In: IEEE International Conference on Robotics and Automation, pp. 3415–3420. IEEE Press (2005)
10. Krebs, A., Pradalier, C., Siegwart, R.: Adaptive rover behavior based on online empirical evaluation: Rover-terrain interaction and near-to-far learning. J. Field Robot. **27**, 158–180 (2010)
11. Ken, H., Peynot, T., Sukkarieh, S.: A near-to-far non-parametric learning approach for estimating traversability in deformable terrain. In: Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 2827–2833. IEEE Press (2013)
12. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge (2000)
13. Costa, A.F., Humpire-Mamani, G., Traina, A.J.M.: An efficient algorithm for fractal analysis of textures. In: 25th Conference on Graphics, Patterns and Images, pp. 39–46. IEEE Press (2012)
14. Liao, P., Chen, T., Chung, P.: A fast algorithm for multilevel thresholding. J. Inf. Sci. Eng. **17**, 713–727 (2001)
15. Rasmussen, C.E., Williams, C.K.I.: Gaussian Processes for Machine Learning. The MIT Press, Cambridge (2006)