Rigidity-Based Surface Recognition for a Domestic Legged Robot

Csaba Kertész, *Member, IEEE*

***Abstract*—Although the infrared (IR) range and motor force sensors have been rarely applied to the surface recognition of mobile robots, they are fused in this paper with accelerometer and ground contact force sensors to distinguish six indoor surface types. Their sensor values are affected by the crawling gait period, therefore, certain components of the fast Fourier transform over these data are included in the feature vectors as well as remarkable discriminative power is observed for the same scalar statistics of different sensing modalities. The machine learning aspects are analyzed with random forests (RF) because of their stable performance and some inherent, beneficial properties for the model development process. The robustness is evaluated with unseen data after the model accuracy is estimated with cross-validation (CV), and regardless whether a Sony ERS-7 walks barefoot or wears socks, the forests achieve 94% accuracy. This result outperforms the state of the art techniques for indoor surfaces in the literature and the classification execution is real-time on the robot. The above mentioned model development process with RF is documented to create new models for other robots more quickly and efficiently.**

***Index Terms*—Surface Recognition; Random Forests; AIBO**

I. Introduction And Related Work

THE wheels are better options on even surfaces while the legged robots traverse on more difficult terrains. With knowledge about the underlying surface, a legged robot can switch to an efficient gait or adapt its walk speed for optimal locomotion. This paper focuses on the context awareness of domestic robots by predicting six surfaces based with built-in sensors of a Sony ERS-7. Besides the focus on the indoor setting and evaluating the machine learning aspects of past researches, the literature review touches the other conditions in outdoor environments and the use of vision sensors.

The surface recognition is less challenging for outdoor robots because the vibration-based solutions perform better with higher irregularities while the indoor floorings challenge the image classifiers with wider variety of colors and textures. Learning visual cues in a house can enhance a vibration model, but creating a generic texture or color based classifier for all kinds of carpets, tiles and other floorings is an overwhelming task. By these reasons, the terrains were detected with higher accuracies by fused modalities outside compared to the indoor floorings [5] and the vision models suit better in natural environments [4, 17]. These experimental conditions are examined in this paper.

Ojeda and Borenstein [11] experimented with a four-wheeled Pioneer 2-AT on six outdoor terrains and different sensors were explored during the classification process with one training and one testing set. Their neural network produced reasonable performance for the inertial sensors (82.7% accuracy), but the cross-validation was not applicable with their small sample set.

Hoepflinger et al [6] extracted the features from joint motor currents and ground contact force measurements to estimate different terrain shapes and surface properties. A robot leg was fixed to a table in their testbed thus the sensor readings were not affected by robot body oscillations. The model performance of their AdaBoost classifier was not estimated with cross-validation and there is a high chance that these models were overfitted.

By fusing tactile, depth sensors and camera, a six-legged walking robot [15] recognized 12 surfaces with a success rate of 95%. Since only one testing and one training set were evaluated, cross-validation was not performed to estimate the model performance. The feature vector size (174) was much higher than the sample set size (84), therefore, the real accuracy of this method is uncertain with possible overfitting.

Unlike the previous examples, several researches executed appropriate estimations about the model performance with k-fold CV. Hoffman et al [8] explored the terrain discrimination with inertial, tactile, and proprioceptive sensors of a crawling legged robot. Two classifiers (support vector machine (SVM), naïve Bayes) were cross-validated and the performance was estimated at 96.3% with four surfaces (plastic foil, cardboard, Styrofoam and rubber). This result can be compared to the previous work of the author [10] where 93% accuracy was estimated with 10-fold CV when a Sony ERS-7 walked on five surfaces (wood, short carpet, carpet, foam mats, vinyl).

The Sony robots were equipped with noisy, low-end accelerometers (120 Hz) while high-end devices were used in some earlier studies with a sampling rate of 44.1 kHz in [2] and 4 kHz in [5]. A C4.5 decision tree was implemented for AIBO in [13] where the single and pair-joint variances of the three accelerometer dimensions (x, y, z, x-y, y-z, x-z) composed the feature vector and a large sample database was collected. The accuracy of 84.9% was estimated for a model of three surfaces (cement, field, carpet) by 10-fold CV.

OctoRoACH went on three surfaces in [1] and the feature vectors were extracted from inertial measurement unit and force sensor readings. Bermudez et al studied the shortest

Manuscript received: August, 31, 2015; Revised November, 14, 2015; Accepted January, 4, 2016.

This paper was recommended for publication by Francois Chaumette upon evaluation of the Associate Editor and Reviewers' comments. This work is supported by the competence development bonuses of Vincit Oy.

The author is with University of Tampere and Vincit Oy, Tampere, Finland, csaba.kertesz@ieee.org.

Digital Object Identifier (DOI): see top of this page.

Fig. 1. The images show the following indoor surfaces from top-left: a) Wood flooring and a short carpet, b) Wood flooring and a soft carpet, c) Normal vinyl, d) 8x8cm porcelain tiles, e) Carpeted floor, f) Slippery vinyl, g) 2x2cm porcelain tiles, h) Laminate flooring and a shag carpet.

sampling window (350 msec) to maximize the classifier performance for a running gait. The SVM hyperparameters were optimized by 10-fold CV on a training set and 93.8% accuracy was estimated.

The model accuracy is calculated properly with an evaluation step of unseen data, like in the following works. Degrave et al [3] researched different sensing modalities for supervised and unsupervised classification with five surfaces (blue foil, styrofoam, linoleum, cardboard and rubber). Their experiments found that non-linear sensor fusion and classifier are necessary for good results. The reservoir computing model of tactile and proprioceptive sensors achieved 84.69% accuracy with a validation set.

Tick et al [12] run a wheeled robot on five surfaces (tiled linoleum, ceramic tiles A/B, short carpeted floor and terrazzo). After many statistics were extracted from accelerometer and angular velocity measurements, they studied the sequential forward floating feature selection to build feature vectors for a Linear Bayes classifier. The cross-validation was not executed in this research, however, the built model was tested under practical conditions and the robot achieved 89% accuracy on unseen sensor recordings.

The reviewed literature solved the surface recognition challenges with machine learning methods, but Holmstrom et al applied experience based models to the problem [9]. Specific gaits were evolved by a genetic algorithm for a Sony AIBO robot to walk on plywood board, thin foam, short carpet and shag carpet. After they found some optimal model parameters empirically, tap-delay Adaline neural networks modeled the experience of the leg joints for every gait-surface combination and the robot predicted a surface according to the actual gait. Though a limited sample set was collected, their initial experiments showed promising results and the model accuracy was estimated to 92% with 5-fold cross-validation.

Various locomotion options and sensor combinations have been examined in the past works: wheeled [11, 12, 16, 17], quadruped [3, 4, 8, 9, 10, 13], six-legged robots [15]. Although some studies analyzed a few sensors [9, 11] or the body oscillations [18], this paper does not go into this direction, but the focus is on the classifier model building and the evaluation to provide a better interpretation of the past results from machine learning point of view. Most reviewed studies evaluated their models properly [1, 3, 8, 9, 10, 12, 13], but some did not employed (cross-)validation to get a better estimate of the real accuracy [6, 11, 15]. This paper uses a four-legged Sony ERS-7 robot and the feature vector is composed by the readings of four sensors to train the models. Nine classifiers are examined and random forests are chosen as they have some unique properties for the further analysis. After the features are extracted with fast Fourier transform and statistics, the model accuracy is estimated with 10-fold CV on a training set, and finally, RF is evaluated on a validation set.

II. Experimental Setup

A. Gait and Surfaces

A Sony ERS-7 walked with singular crawling during the experiments to protect the weak servomotors and maintain the stability by having three legs always on the ground [10]. The walk period (φ) was 2400 msec (2.4 Hz) and the speed was around 2cm/sec. The robot traversed on six different, common surface types found in households (Fig. 1):

- 8x8cm porcelain *tiles* and 2x2cm porcelain tiles.
- Lacquer coated *wood flooring* and laminate flooring.
- A bit less rigid normal and slippery *vinyl flooring*.
- *Carpeted floors* with 0.5-1mm thick plastic foam.
- 2-3mm thick *short carpets*.
- 13mm thick *soft carpet* and shag carpet.

Unlike when the robots walked on one example per domestic surface type in [5, 8, 9, 10, 13, 15], AIBO run on multiple examples per type to gather samples in a more generic manner in this paper. The intraclass variability in the dataset were higher in this way and the interclass correlation had a higher chance tough these properties were not studied. Although the body oscillations were mainly influenced by the rigidity and the slipperiness, the surfaces were classified here by the first criterion; for example, soft and shag carpets were placed in the same class. To make the problem even more challenging, the samples were collected with socks and walking barefoot. (The robot worn dog socks with different anti-slip patterns in the experiments, their exact installation can be found in [10].) The idea behind the mixed usage of socks was invented when the author did not find any significant effect on the recognition accuracy with two classifiers trained with no socks/socks cases separately and when the samples were merged together in a single classifier during the initial experimentation. On one hand, the complexity was reduced with collapsed samples, on the other hand, half of the AIBO owners draw anti-slip socks on their robots and the remaining do not use socks at all [10]. It is a clear advantage to use one machine learning model regardless of the owner preference.

B. *Sensors*

The tactile sensors have been widely used for surface classification [1, 2, 3, 5, 8, 10, 11, 12, 13, 15] to measure the body oscillations during locomotion, and in AIBO, there is a low-cost accelerometer in the torso with a 120 Hz sampling rate.

To the best knowledge of the author, the infrared range sensors were considered by Ojeda and Borenstein [11] for terrain discrimination of mobile robots in the past and they proposed the frequency domain components of these sensors to complement other devices. A built-in IR sensor with a 25 Hz sampling rate on the chest of AIBO was utilized in this paper which was directed to the flooring by 30 degrees.

The advanced force sensors have been often attached to the tip of the robot legs to measure the ground contact forces [3, 6, 8, 15], but the paws in this experiment had a simple two-state contact force sensor with a 10 Hz sampling rate. While proprioceptive sensors (e.g potentiometers) were researched in [1, 3, 6, 8, 9], the ERS-7 model has a force sensor in each leg joint. The previously mentioned papers combined all joints in the sensor readings, however, the author found real discriminative power for the hip joints of the hind legs during the initial experiments. After these hip joints were included in the feature extraction, any other joint had no influence on the classifier performance.

C. *Sample Collection*

The initial body oscillations (first two walk periods) have an undesired effect on the extracted features as noise, therefore, they were omitted. In [14], the duration of the first step was excluded with similar purpose for a hexapod.

The sampling frequency of the operating system in AIBO is 31.25Hz and the crawl gait is slow, therefore, two walk periods (4.8 seconds) of sensor data were used to extract the feature vectors. This sliding window size is enough to contain a full walk period all the time to catch all body oscillations relevant to the current surface. The author varied the window size in the initial experiments, but the shorter length increased the classifier complexity (random forest size) without any gain in the accuracy. Similar to the this size, Hoffman et al [7] found the 6 seconds-long sensor readings the most accurate with their four-legged robot and an other work [5] concluded the 4 seconds-long window over 1 second. However, shorter windows can suit better for different robots or gaits; Bermudez et al [1] found a 350 msec time window enough for a running hexapod robot to maximize the model accuracy.

30709 samples were collected for the dataset what the author opened under the name of Indoor Surface Recognition Dataset (ISRD - DOI: 10.13140/RG.2.1.3877.5764). The corpus was split into a training (S_T) and a validation set (S_V) randomly in 40%/60% partitions. The training set had 2026 wood flooring, 2109 vinyl, 2214 tiles, 1402 carpeted floor, 2510 short carpet and 2112 soft carpet samples, balancing all classes around 2000 samples. The validation set had 2798 wood flooring, 2809 vinyl, 3128 tiles, 1970 carpeted floor, 2276 short carpet and 5355 soft carpet samples. The first role of S_T was to estimate the classifier accuracy with cross-validation and the second was to build the final models for the evaluation of S_V. Such a large validation set has not been reported in the literature, 75%/25% split was defined in [1] and 84%/16% in [12]. Note that the less fraction of the samples are included for training the more difficult for a classifier to predict the validation set.

III. FEATURE VECTOR

Before the classifiers are trained, a feature vector must be defined. This chapter describes how the features were extracted from the sensor data streams.

The feature vectors of surface models contain spectral components and statistical descriptors which are computed from a time window over the raw sensor data. The feature vector size (48) in this paper is on the average compared to previous works:

- Vail and Veloso [13] used 6 features derived from the accelerometer data.
- Bermudez et al [1] suggested 15 statistical features.
- Hoepflinger et al [6] defined 20 features from motor currents and ground contact forces.
- Tick et al [12] selected 68 features out of 864.
- Weiss et al [16] generated 128 FFT components from accelerometer data.
- Walas [15] had 174 features generated from tactile, depth sensors and camera.

The feature extraction is a crucial part of the classification process because the models must comprehend discriminative features to provide good predictions. In the literature, either the researchers selected some scalar statistics without deeper analysis [13, 15] or many features were generated in order to run through feature selection. These practices can lead to the usage of non-relevant features in the first case or ending up different optimal statistics for every sensor in the second case. The author of this paper considered many statistics for each sensor and the best were selected manually after an iterative examination of the feature importances in the initial experimentation. The automated feature selection ended up with various statistics for tactile sensors in [12] however the same were optimal for all sensing modalities in this work.

Albeit the statistical moments have lower computational costs than fast Fourier transform (FFT) analysis and they were preferred in [1], the experiments in this paper did not find these moments sufficient to distinguish the surfaces without the FFT magnitudes. The author believe that Bermudez et al found the moments suitable as their surfaces were very distinct.

Some past works used all components of the Fourier transform [11, 16], some reduced the dimensions with principal component analysis [2] or similar method. Holmstrom et al [9] calculated the FFT on the time series of the proprioceptive sensors in AIBO and the third harmonic peak (~4.5Hz) showed significant difference for multiple surfaces. The F\ourier analysis of tactile sensors in a hexapod robot [15] showed varying magnitudes for more surfaces below 9 Hz and the frequency range 0-4Hz contained most differences, similar to [9]. The author of this paper found that the useful frequency bands were in relation to the walk period, namely, its overtones and the inharmonic partials ($k*\varphi$) hold most information for surface classification where $k \in \{1/16, 1/8, 1/4, 1/2, 1, 3/2, 2\}$. These frequencies were confirmed with feature selection when the first 20 FFT amplitudes of several sensors were added to the feature vector during the initial experiments and the same bands had remarkable feature importances while other bands had negligible. This finding needs a detailed theoretical analysis in the future, but it is not part of the current study.

Every sensor had 150 measurements in two walk periods and the feature vectors were computed over this time window. Six frequency bands contained the proposed overtones and inharmonic partials in the following FFT components: 1^{st}, 2^{nd}, 3^{rd}, 6^{th}, 9^{th} and 12^{th}. Although these bands had good discriminative power for one sensor, but after the same bands were added to the feature vector for new sensing modalities, the gained overall improvements had a decreasing trend. While all six components were worth for the accelerometer (z-axis), the IR sensor had five and the force sensors three. This result may suggest that the FFT analysis of different modalities capture similar discriminative capability caused by the body oscillations from the classifier point of view which implies these oscillations as main influence on the IR sensor readings unlike the surface reflections. This phenomena requires further analysis as well.

A. *Accelerometer Sensor*

Median, maximum, skewness and root mean square (RMS) amplitude were computed over the sliding window of the accelerometer angles (x, y, z). The robot walk on rigid flooring produces vertical body oscillations, which can be detected in the z dimension [16, 18], while soft surfaces absorb these anomalies. The time series from z axis were transformed to the frequency domain by FFT and the six proposed components were added to the feature vector. 18 features were generated in overall from the accelerometer.

B. *Infrared Range Sensor*

The IR range sensor on the chest operates within [10; 90 cm] and the robot body oscillations alter these values. The interquartile range (IQR), maximum, skewness, RMS amplitude statistics, the first five proposed and the largest FFT components were added to the feature vector. 10 features were originated from this sensor.

C. *Leg Force Sensors*

The force sensors in the hip joints of the hind legs were chosen (see Chapter II.B) for feature extraction. The same statistics (IQR, maximum, skewness, RMS amplitude) were calculated again along with the first three proposed and the largest FFT amplitudes. These sensors contributed 16 features.

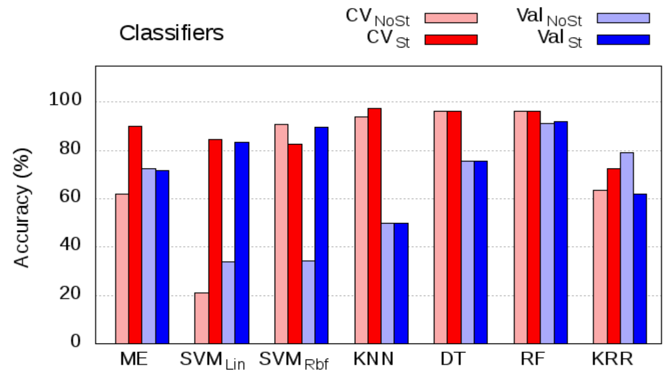

Fig. 2. Every first two bars (CV_{NoSt}, CV_{St}) show the classifier performances in cross-validation and latter two (Val_{NoSt}, Val_{St}) for unseen data. Feature standardization was employed for the even bars and there was no feature preprocessing in the odd bars. Classifiers: maximum entropy (ME), support vector machine with linear kernel (SVM_{Lin}), support vector machine with radial basis function (SVM_{Rbf}), k-nearest neighbor (KNN), decision tree (DT), random forest (RF) and kernel ridge regression (KRR).

D. *Ground Contact Force Sensors*

The four paws of the Sony ERS-7 are two-state buttons. They are pressed more likely when the robot walks on a rigid surface compared to a carpet. Therefore, the pressed durations in a walk period provide a good metric about the rigidity. A simple sum was calculated from these sensors which produced the last four values to the feature vector.

IV. Classifier Selection

According to the well-known no free lunch theorem in machine learning, there is no universal “best” classifier for all problems in the world and different methods can achieve similar, satisfactory results. Nine classifiers were compared in this study (Fig. 2) and the effects of feature standardization were examined during cross-validation and validation phases. (The classifier hyperparameters can be looked up in [10].) The relevance vector machines and naïve Bayes (NB) classifiers achieved low accuracies (< 25%) thus they were omitted in Fig. 2. The latter was unexpected as the Bayes classifiers has been performed remarkable in the literature [8, 10, 12, 16].

Similar to a previous work of the author [10], Weiss et al studied several algorithms with 10-fold cross-validation [16]. Although the decision tree (J4.8 variant) did not yield good results, SVM, KNN and NB had the highest model accuracy estimations in [16]. The author found that NB, SVM and DT were ahead of KNN in [10], but in this paper, SVM, DT, RF, KNN and ME were over 80% in the CV phase.

SVM classifiers are sensitive to the missing feature standardization what is reflected on Fig. 2. SVMs had improved accuracies over 80% in almost every case when the data was standardized. On the other hand, the feature preprocessing caused performance losses in some situations (CV of SVM_{RBF}, validation of *KRR*). Although the SVM classifiers had reasonable accuracies in the validation phases, the random forest delivered the most stable performances (91-96%), regardless of applied or absent feature preprocessing. Other classifiers had varying, lower results.

Though neural networks [3, 5, 11, 16], SVMs [1, 8, 10, 15, 16] and decision trees [10, 13, 16] have broad literature in surface classification, the random forests have not been examined at all. As a consequence of the missing experiments, the potentials of this classifier family have not been exploited because the decision trees have limited learning capabilities compared to RF. The random forests were chosen for the further experiments in this paper to investigate the uncovered topics and benefit from the built-in variable importance measures of RF for feature ranking.

A common practice in machine learning to remove the outliers from the database as they can confuse the classifiers, but the author did not find any impact on the accuracy in the initial experiments hence they were left.

Usually it is not expressed in the robotics papers, but the CV results of a classifier can not be matched directly to the model accuracies calculated on unseen data since the first gives only an estimation about the latter. For example, there was $CV_{St} >> Val_{St}$ for ME, DT and KNN (Fig. 2). Therefore, the cross-validation results are compared to the CV values of the literature in Chapter V.A and the model accuracy (Chapter V.B) is presented against the relevant works.

V. Model Analysis with Random Forest

The random forests have been implemented for many problems except surface recognition. This paper fills this gap by examining the RF models closely as previous works have not been went into details about the effects of classifier hyperparameters [1, 5, 6, 8, 10, 13, 16], only the feature selections have been researched [3, 11, 12, 15].

The forest dimensions depend on the maximum tree depth and the forest size. $rf_{x,y}$ defines a random forest with maximum tree depth x and forest size y. The minimum sample count on a leaf for splitting is an other important parameter and it is recommended to set around $|S_T|$ / 100. It was fixed to 100 in these experiments to avoid overfitting.

A. Model Accuracy Estimation

The k-fold cross-validation does not replace the model verification on unseen data, but it gives a reasonable estimate about the model performance. 10-fold cross-validation was run with an $rf_{20,20}$ on the training set to estimate the model performance in this paper and 96.2% accuracy was achieved with six classes. The surface recognition gets challenging by increasing the surfaces as the model must distinguish more and more classes correctly. This novel approach outperformed other methods in the literature because the performance is similar or higher than the estimations of less indoor surfaces and it was better from [5] by 5.7%:

- 3 surfaces: 84.9% in [13].
- 4 surfaces: 92% in [9], 96.3% in [8].
- 5 surfaces: 93% in [10], 96.2% in [3].
- 6 surfaces: 90.5% in [5].

Note that the evaluation above excludes the earlier studies for outdoor terrains [1, 2, 11, 17] since the surface recognition must distinguish more subtle details in the body oscillations on domestic floorings (see in Chapter I) with less surface irregularities. [6, 12, 15] were also omitted by the reason of the missing cross-validation step.

B. Model Accuracy

The cross-validation estimates the model accuracy to some extent, but the results for KNN and DT were far from the real performances in Fig. 2 what warns about the limitations. The models must be built with training samples and tested on unseen data to get a proper measure hence the random forests in this subchapter were constructed with S_T and evaluated with S_V (see in Chapter II.C).

Depending on the random forest parameters, a model can underfit the data if the forest is too small or overfitting happens when too large. Fig. 3 represents how $rf_{i,j}$ models ($i,j \in [4, \ldots, 80]$) were evaluated for their accuracy and memory usage in the function of the maximum forest size and tree depth. The blue-green area on Fig. 3.a shows the underfitting models and the red-orange-green area on Fig. 3.b contains the overfitting models. The area of ($i \in [10,\ldots, 80]$; $j \in [20,\ldots, 30]$)

TABLE I. Confusion Matrix of an $RF_{20,20}$ Model
The rows show the real surfaces and the columns how they were classified.

%	W	SC	C	V	T	CF
W	**91.57**	0.61	0.54	2.68	4.50	0.11
SC	0.70	**90.99**	3.21	1.49	2.02	1.58
C	0.15	2.22	**90.66**	0.77	2.11	4.09
V	4.13	3.88	0.61	**87.86**	2.85	0.68
T	0.67	0.93	1.28	2.11	**94.66**	0.35
CF	0.36	0.30	1.07	0.86	0.61	**96.80**

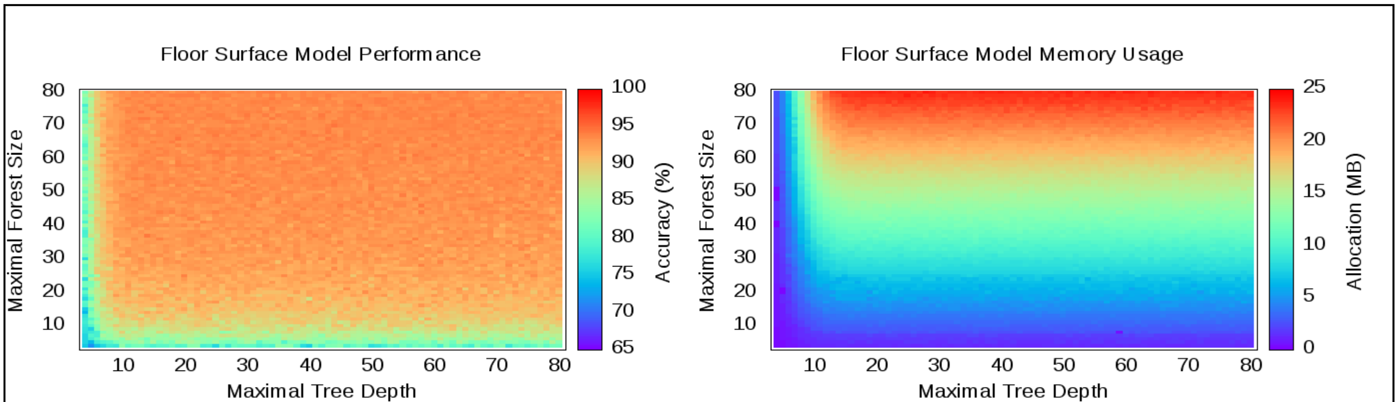

Fig. 3. The diagrams show the random forest model accuracies (3.a) and memory usages (3.b) in the function of the maximum tree depths and forest sizes.

TABLE II. FEATURE IMPORTANCES
THE ODD COLUMNS CONTAIN THE FEATURES, THE EVEN COLUMNS SHOW THE RELATIVE IMPORTANCES FROM H_2O MACHINE SOFTWARE (SCALE: 10^2) IN DECREASING ORDER.
(ACCELEROMETER: A_x, A_y, A_z; INFRARED RANGE SENSOR: IR; FORCE SENSORS OF HIP JOINTS IN HIND LEGS: F_{LH}, F_{RH}; GROUND CONTACT FORCE SENSORS: GF_{LF}, GF_{LH}, GF_{RF}, GF_{RH})

iqr(F_{LH})	191	fft_{max}(F_{RH})	116	skew(F_{RH})	56	rmsa(A_z)	27
rmsa(F_{RH})	147	sum(GF_{RH})	111	fft_3(F_{LH})	56	fft_9(A_z)	24
med(A_y)	147	max(F_{LH})	107	skew(IR)	55	fft_2(A_z)	24
fft_1(IR)	146	fft_5(IR)	99	iqr(F_{RH})	53	fft_{12}(A_z)	23
sum(GF_{LF})	146	max(A_y)	93	max(A_x)	52	fft_1(A_z)	22
rmsa(A_x)	141	rmsa(F_{LH})	75	fft_3(F_{RH})	51	iqr(IR)	22
sum(GF_{LH})	135	fft_1(F_{LH})	74	skew(A_x)	48	fft_3(A_z)	22
sum(GF_{RF})	125	skew(F_{LH})	72	max(A_z)	33	fft_6(A_z)	22
fft_1(F_{RH})	124	max(IR)	61	fft_3(IR)	33	fft_4(IR)	21
fft_{max}(IR)	124	skew(A_y)	60	max(F_{RH})	30	fft_2(F_{LH})	20
rmsa(IR)	123	fft_{max}(F_{LH})	59	fft_2(IR)	30	fft_2(F_{RH})	17
med(A_x)	123	rmsa(A_y)	58	skew(A_z)	28	med(A_z)	6

TABLE III. ACCURACIES OF VARIOUS SENSING MODALITIES
(ACCELEROMETER: A; GROUND CONTACT FORCE SENSORS: GCF; FORCE SENSORS OF HIP JOINTS IN HIND LEGS: F, INFRARED RANGE SENSOR: IR)

A	GCF	F	IR	A+GCF	A+GCF+F	All
48.1	62.7	69.7	55.2	75.8	87.4	92.0

in Fig. 3.a and Fig. 3.b provides an accuracy plateau of 91-94%. The Table I shows the confusion matrix of an $rf_{20,20}$ model (accuracy: 92.09%, precision: 92.31%) and most misclassifications (orange) happened between classes with similar rigidity. The tiles and carpeted floor had the highest accuracies, therefore, they caused unique body oscillations.

Two earlier works reported model accuracies for five indoor surfaces. Degrave et al [3] achieved 84.69% and Tick et al [12] 89%. The new method of this paper realized a notable 94% for six surfaces and the model generalized over a mixed set of barefoot and sock samples while each class contained several surface examples. This result outperforms [3] and [12] with more surfaces, generalization power and higher accuracy.

C. Feature Importances of Different Sensing Modalities

The variable importances describe the discriminative contributions of the individual features to the model accuracy. A reason for choosing the random forests was the inherent capability to calculate these values after the training phase as Table II shows for an $rf_{20,20}$ model. The author experienced that a feature vector contained unnecessary weak predictors if 5+ features had their relative importances below 10 and removing such variables did not effect the model accuracy. All features in Table II had significant discriminative ability, they differed only in the relative importances to each other.

It was interesting that every modality was significant on average, but the z-axis of the accelerometer produced relative weak discrimination compared to other axes and sensors, its features had low ranks. This result was against the expectation that this axis contains the most descriptive components of the body oscillations [16, 18].

The maximum and 3rd statistical momentum over multiple sensors are in the middle columns of Table II, they provide a stable, average discrimination. The new RMS amplitude statistics (blue) have good performance among the other features and the ground force sensors (yellow) are outstanding despite they are simple two-state sensors. The best FFT features (orange) are principal and maximal coefficients, generated by the infrared range and the motor force sensors. These sensors have been underutilized in the surface recognition, Ojeda et al proposed the infrared range sensors as complementary for inertial sensors [11] what could be originated in more irregularities of the outdoor surfaces. Bermudez et al attached the force sensor to deformable polymeric legs [1] while the force sensors are placed in each leg joints of AIBO.

Table III shows the $rf_{20,20}$ model accuracy when different feature subsets were used for training and validation. The accelerometer (18 features) had the lowest score among the individual sensor features and the leg based features (GCF, F) were strong, similar to [3], while proprioceptive and feet pressure features had higher rankings over the accelerometer in [8]. The accelerometer has been the most popular sensor for surface recognition, but this modality had the lowest relative discriminative power in this paper and in [3, 8]. GCF sensors had good results again, similar to Table II, although they produced only four features. These latter sensors added the biggest contribution to the sensor fusion (A+GCF - 27.7%), force sensors 11.6% (A+GCF+F) and IR 4.1%. The author examined the joint angle sensors in the initial experiments as well, but that modality did not improve the RF models hence they were not included in this work.

D. Computational Requirements

The random forests were coded in C++ with the OpenCV library. A model was built in 1-4 seconds on a first generation Core i7 (1.86Ghz) depending on the forest size. This result outperformed all training times in [16], considering the weaker processor and the larger training set (12373 vs. 9203 samples in [16]) of this paper. The feature extraction with three FFT analyses took 3 msec on a MIPS CPU (576 Mhz) in AIBO and a smaller forest ($rf_{7,5}$) with 90.9% accuracy was selected because of the trade-offs in embedded platforms. This smaller RF predicted a surface in 20-90 ìsec with 833KB RAM.

VI. SURFACE MODEL DESIGN

After the author worked on the machine learning problem of surface recognition and reviewed the available literature, some advices can be given for future researches. These experiences were gathered with a quadruped Sony AIBO, but they may be applicable for other robots:

1. The FFT amplitudes have good discriminative power for sensors of different sensing modalities. The FFT components with the overtones and the inharmonic partials of the walk period are advised (see Chapter III), but after the first sensor is added to the feature vector, only the lower partials of the other sensors contribute improvements to the model performance.

2. Ground force sensors (even simple ones) predict the surface rigidity very well (see Chapter V.C).

3. Recommended statistics for feature extraction: RMS amplitude, IQR, median, skewness and maximum.

4. The applied machine learning is rather a "black art" than exact science nowadays. The author proposes the random forests for the initial experiments and feature selection, but RFs are not the ultimate answer for the surface classification as (legged) robots with different dynamics, sensors or modified feature extraction may need other optimal method. (Usually, powerful features produce similar accuracy with more classifiers.) Although the author selected the features manually, but he believes that the variable importance functions of RF are beneficial to execute this process semi-automatically. The feature vector can be initialized with the FFT amplitudes of a sensor and new features/statistics can be added to the vector with sequential floating forward selection, similar to [12], based on the feature importances. The two main discrete parameters (maximum forest size and tree depth) of RFs are an advantage to control the forest size while many machine learning algorithms have several float hyperparameters. To give an insight to the RF properties, figures can be drawn (Fig. 3) to visualize the sweat spots where these parameters have reasonable accuracy without under- and overfitting. Note that bigger sample sets need longer time to produce these diagrams, up to several hours.

VII. Conclusions

The paper detailed the creation of a random forest model to recognize six indoor surface types. Although the random forests have not been evaluated for surface recognition in the past, they were cross-validated to estimate the model accuracy and the real performance was computed with unseen data in this study. Both results (cross-validation – 96.2%, accuracy - 94%) outperformed the state of the art researches for domestic environments despite the smaller training set, the intraclass variability and the mixed barefoot/socks recordings in the sample sets. The new method had low computational and memory requirements to run the model in real-time on a Sony AIBO. The author found some useful practices what can be applied to the surface model development (see Chapter VI) in the future. Especially, the random forest classifier has some inherent properties how this process can be more effective.

Other contributions were the successful sensor fusion of some underutilized sensors (infrared range, motor force) in the field. A few FFT amplitudes were proposed for surface recognition whose bands were determinated by the overtones and the inharmonic partials of the crawling walk period. The feature selection confirmed the importance of these magnitudes hence future researches with legged robots are encouraged to use these frequencies. It was also found that the classifier captured similar prediction capabilities of the same FFT components of different modalities which rooted on the body oscillations caused by the walk period.

Many statistics were explored to find the maximum, skewness, interquartile range and median beneficial for more sensors. The root mean square amplitude was applied efficiently to all modalities though it has not been considered for surface classification earlier.

Future work can consider more surface types, the examination of traversing the surface edges and the detection of slippery surfaces although there are several limitations in this paper. One gait at a fixed speed was analyzed thus more experiments must be executed with varied conditions since several past studies focused on multiple gaits [8, 9, 12] and speeds [10, 16]. The infrared range sensor was directed to the ground in this paper and the effects of small objects (e.g LEGO bricks) on the floor have not been researched yet.

Acknowledgment

Thanks for the support from my company and the patience of my family when the samples were collected.

References

[1] F. L. G. Bermudez, R. C. Julian, D. W. Haldane, P. Abbeel, and R. S. Fearing, "Performance Analysis and Terrain Classification for a Legged Robot over Rough Terrain," Int. Conf. on Intelligent Robots and Systems (IROS), pp. 513-519, 2012.

[2] C. A. Brooks and K. Iagnemma, "Vibration-based Terrain Classification for Planetary Exploration Rovers," IEEE Transactions on Robotics, Vol. 21(6), pp. 1185–1191, 2005.

[3] J. Degrave, R. V. Cauwenbergh, F. Wyffels, T. Waegeman, and B. Schrauwen, "Terrain Classification for a Quadruped Robot," Int. Conf. on Machine Learning and Applications (ICMLA), 2013.

[4] P. Filitchkin and K. Byl, "Feature-Based Terrain Classification For LittleDog," IEEE IROS, pp. 1387–1392, 2012.

[5] P. Giguere and G. Dudek, "A Simple Tactile Probe for Surface Identification by Mobile Robots," IEEE Transactions on Robotics, Vol. 27(3), pp. 534-544, 2011.

[6] M. A. Hoepflinger, C. D. Remy, M. Hutter, L. Spinello, and R. Siegwart, "Haptic Terrain Classification for Legged Robots," Int. Conf. on Robotics and Automation (ICRA), 2010.

[7] M. Hoffmann, N. Schmidt, R. Pfeifer, A. Engel, and A. Maye, "Using Sensorimotor Contingencies for Terrain Discrimination and Adaptive Walking Behavior in the Quadruped Robot Puppy," Int. Conf. Simulation of Adaptive Behaviour (SAB), pp. 54-64, 2012.

[8] M. Hoffmann, K. Štěpánová, and M. Reinstein, "The Effect of Motor Action and Different Sensory Modalities on Terrain Classification in Quadruped Robot Running with Multiple Gaits," J. of Robotics and Autonomous Systems, Vol. 62(12), pp. 1790-1798, 2014.

[9] L. Holmstrom, A. Toland, and G. Lendaris, "Experience Based Surface Discernment by a Quadruped Robot," Symp. on Computational Intelligence in Image and Signal Processing (IEEE-CIISP 2007), pp. 409-414, 2007.

[10] C. Kertész, "Exploring Surface Detection for a Quadruped Robot in Households," 14th IEEE Int. Conf. on Autonomous Robot Systems and Competitions (IEEE-ICARSC), 2014.

[11] L. Ojeda, J. Borenstein, G. Witus, and R. Karlsen, "Terrain Characterization and Classification with a Mobile Robot," J. of Field Robotics, Vol. 23(2), pp. 103–122, 2006.

[12] D. Tick, T. Rahman, C. Busso, and N. Gans, "Indoor Robotic Terrain Classification via Angular Velocity based Hierarchical Classifier Selection," Int. Conf. on Robotics and Automation (ICRA), 2012.

[13] D. Vail and M. Veloso, "Learning from Accelerometer Data on a Legged Robot," 5th IFAC/EURON Symposium on Intelligent Autonomous Vehicles, 2004.

[14] K. Walas, "Tactile Sensing for Ground Classification," 1st Int. Work. on Perception for Mobile Robots Autonomy (PEMRA), 2012.

[15] K. Walas, "Terrain Classification and Negotiation with a Walking Robot," J. of Intelligent & Robotic Systems, Vol. 78(3-4), pp. 401-423, 2015.

[16] C. Weiss, N. Fechner, M. Stark, and A. Zell, "Comparison of Different Approaches to Vibration-based Terrain Classification," 3rd European Conference on Mobile Robots (ECMR), pp. 7-12, 2007.

[17] C. Weiss, H. Tamimi, and A. Zell, "A Combination of Vision- and Vibration-based Terrain Classification," Int. Conf. on Intelligent Robots and Systems (IROS), pp. 2204-2209, 2008.

[18] Y. Chen, C. Liu, and Q. Chen, "A Vestibular System Model for Robots and Its Application in Environment Perception," Int. Conf. on Robotics and Automation (ICRA), pp.. 230-235, 2010.