Friction from Vision: A Study of Algorithmic and Human Performance with Consequences for Robot Perception and Teleoperation

Martim Brandão, Kenji Hashimoto and Atsuo Takanishi

***Abstract*— Friction estimation from vision is an important problem for robot locomotion through contact. The problem is challenging due to its dependence on many factors such as material, surface conditions and contact area.**

In this paper we 1) conduct an analysis of image features that correlate with humans' friction judgements; and 2) compare algorithmic to human performance at the task of predicting the coefficient of friction between different surfaces and a robot's foot. The analysis is based on two new datasets which we make publicly available. One is annotated with human judgements of friction, illumination, material and texture; the other is annotated with static coefficient of friction (COF) of a robot's foot and human judgements of friction. We propose and evaluate visual friction prediction methods based on image features, material class and text mining. And finally, we make conclusions regarding the robustness to COF uncertainty which is necessary by control and planning algorithms; the low performance of humans at the task when compared to simple predictors based on material label; and the promising use of text mining to estimate friction from vision.

I. Introduction

The coefficient of friction between robot contacts and the ground is an important input to model-based motion planning and trajectory optimization methods with dynamics. Wrong estimations of friction may lead to slipping, which in turn causes challenges to state-of-the-art planners and controllers. This fact has motivated recent work on state estimation and slippage controllers [1], [2], [3], [4]. Friction prediction from vision is not only an important problem but also a challenging one: friction depends on many factors such as contact area, surface conditions, material and context; and is still not completely understood in humans despite some interesting findings [5], [6], [7], [8].

From a robotics perspective, the existence of this paper is especially motivated by the recent trend in legged and humanoid robot motion planning field towards optimization-based planners that rely explicitly on "known" coefficient of friction values. Furthermore, this paper is motivated by the lack of public datasets for understanding humans' perception of friction. We provide two datasets to the community, from which we take conclusions regarding not only robot perception but also robot teleoperation and humans' visual perception of friction.

OSA+F dataset (8 materials, 96 images)

GTF dataset (14 materials, 43 images)

Fig. 1. The two datasets collected for this paper, one example per material category. OSA+F dataset: carpet/rug, concrete, fabric/cloth, granite/marble, metal, stone, tile and wood. GTF dataset: asphalt, brick, carpet/rug, cobble, concrete, dirt, granite/marble, leaves, linoleum, metal, mud, stone, tile and wood.

The contributions of this paper are the following:

- We describe, analyze and publicly provide[1] two friction-from-vision datasets in Section III. One is targeted at quantification and benchmarking of friction estimation methods for robot locomotion. The other is targeted at understanding visual perception of friction by humans.
- We propose and evaluate methods for friction estimation based on intrinsic images, gradient images, semantic class predictions and text mining through word embeddings (Section IV).
- We make conclusions regarding the error associated with friction estimation from vision, most informative features for prediction and the dangers of assigning the friction task to human teleoperators of robots.

II. Related work

The friction estimation literature in robotics has been mostly focused on its measurement during contact. Examples include COF estimation using specially-designed sensors [9], and material classification through dynamic friction model fitting while stroking surfaces with a robotic finger [8]. In the legged robot locomotion literature, there is an interest in identifying slips when they occur [1], [2], [3], in order to trigger changes in controllers [1] or activate reflexes [4]. For example, [1] estimated slipping force by comparing predicted ground reaction forces and those measured with

*This work was supported by JSPS KAKENHI Grant Number 15J06497 and ImPACT TRC Program of Council for Science, Technology and Innovation (Cabinet Office, Government of Japan).

M. Brandão is with the Graduate School of Advanced Science and Engineering, Waseda University.

K. Hashimoto is with the Waseda Institute for Advanced Study, and is a researcher at the Humanoid Robotics Institute (HRI).

A. Takanishi is with the Department of Modern Mechanical Engineering, Waseda University; and the director of the Humanoid Robotics Institute (HRI), Waseda University.

[1]The datasets are publicly available at:
http://www.martimbrandao.com/friction-from-vision/

a force sensor on the foot. On the other hand, [2] uses Kalman filtering of IMU measurements to detect slippage, and [3] applies a similar approach to a quadruped robot which considers active contact information as well.

While such methods focus on detecting slippage for control purposes, motion planning algorithms have also been adapted to decrease the risk of slipping even in low friction conditions. For instance, [10] proposes a method for grasp synthesis prioritizing low "friction sensitivity", such as to prefer grasp configurations that are stable even for low COF. Similarly in the biped locomotion literature, [11] changes parameters regulating center-of-mass motion such that the minimum COF where the robot can walk without slipping is decreased. And going further [12] plans robot configurations and walking speeds for the same purpose.

Even if planning algorithms can adapt motion to increase the range of COF where the robot can walk, it is still important to have an estimate of the actual friction and its uncertainty. Otherwise, planned motions may be too conservative and suboptimal, or too aggressive considering reflex controllers' robustness. One option to tackle this problem is through learning from experience. Notably, [13] uses visual terrain classification and slope to estimate friction on a rover. The authors train their models on image sequence datasets of rover navigation. Compared to that work, we provide open datasets where algorithms can be benchmarked and human judgement data as well. We also propose a text mining method that can provide a prior for friction even on terrains without previous locomotion experience.

One of the friction estimation methods from text mining we propose in this paper is similar to the method used for affordance estimation in [14]. There, the task is to automatically classify which actions can be applied to different objects, which the authors compute using distances between vector representations of words. While [14] computes noun-verb relationship pairs, we take into account a list of possible words related to slippage and their distance to a material noun. Both our algorithm and others relying on material classification are subjected to errors in image-based material classifier performance. State of the art of material classification is currently at around 70% [15], [16], and progress in the area has been improving quickly. Also, recent GPU-based scene understanding algorithms, for example [17]'s integrated SLAM and scene understanding, are becoming fast enough for robot locomotion applications, thus further motivating this paper.

Human performance might inform the robotics and computer vision community of features to use for prediction. Healthy humans are capable of walking in diverse environments with different degrees of friction, and prospectively adapt walking style before touching slippery surfaces [18]. Still, humans make friction judgement mistakes that lead to slipping for example due to over-reliance on gloss or other lighting-related visual features [5]. Humans also use other cues such as texture smoothness [6], and presence of water or other contaminations [7].

III. Friction From Vision datasets

We now describe our methodology for obtaining the two datasets used for analysis: the OpenSurfaces and Friction dataset (OSA+F) and the Ground-truth coefficient of Friction dataset (GTF).

A. OpenSurfaces and Friction (OSA+F)

The OSA+F dataset is targeted at prediction and understanding of human judgements of surface friction during human locomotion.

Due to its variety of annotations, we started from the open and crowd-sourced OpenSurfaces dataset [19], along with the texture attribute annotations of [16], referred to as OSA (OpenSurfaces plus texture Attributes). Each image is annotated with segments drawn by the subjects and each segment is attributed an object name, material class (1 out of 22) and the applicability of texture classes (boolean vector of size 11, e.g. whether the segment's texture is chequered or not, marbled or not, etc). Albedo and reflectance judgements also exist for most segments. We considered the data available with the OSA dataset most suitable for the friction estimation task since human judgements of friction are usually associated with gloss [5], material and texture [6].

We selected a high-quality, class-balanced subset of the OSA dataset appropriate for our task. First, for high-quality annotations, we discarded segments with negative judgement scores. Since our goal is to obtain a dataset for friction estimation of locomotion surfaces, we only considered segments corresponding to traversable planar surfaces. Traversability was manually annotated by the authors. From the high-quality traversable segments we selected 96 segments for the OSA+F dataset. These were obtained by solving a mixed-integer linear program maximizing total segment area, subject to the constraints: 1) each material has exactly 12 occurrences in the dataset, 2) each texture has at least 10 occurrences in the dataset, 3) each image has only one segment in the dataset (to prevent similar segments from the same image). The resulting OSA+F dataset thus consists of 96 segments, from 96 images, and 8 material classes with 12 occurrences each. We show one example image for each material class in Figure 1.

We collected human judgements of friction for each image segment through an online survey with random image order, one image per page, prepared using the Limesurvey software [20]. Subjects were 14 graduate students from the mechanical engineering department with normal or corrected-to-normal visual acuity. Each image segment was judged by the subjects using a slipperiness Likert scale of 1 to 6 (i.e. 1 least slippery, 6 most slippery). We opted for this scale after preliminary experiments showing larger scales to be difficult to judge, "slipperiness" to be easier to rate than "friction", and because the same scale is used on different material judgement experiments in the human vision literature [21]. The explanation of the scale was present in all pages. The questions were framed as how slippery the subjects expected the surfaces to be in case they were walking on them with

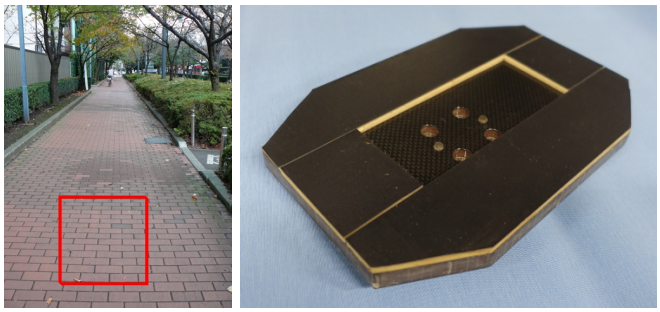

Fig. 2. Left: example image from the GTF dataset overlaid with a red square indicating where the coefficient of friction was measured. Right: sole of the humanoid robot foot used for the coefficient of friction experiments, GTF dataset.

their normal shoes. As a post-processing stage we normalized judgements to a friction scale instead of slipperiness (i.e. $y = 1 - \frac{y_{\text{likert}}}{6}$, thus 0 is lowest friction, $\frac{5}{6}$ highest). On the survey, segments were indicated by a red square overlaid on the image, computed as the largest-area square inside the OSA segment. See Figure 2 (left) for an example.

B. Ground-truth coefficient of Friction (GTF)

The GTF dataset is targeted at prediction of coefficient of friction estimates from images for robot locomotion, and includes also human predictions as a baseline for algorithm evaluation. Importantly, coefficient of friction depends on properties of both surfaces in contact, and thus the main objective of building this dataset is not to train predictors applicable to all robots, but to quantify humans' and algorithms' performance at the task. Our assumption is that the conclusions taken from our robot foot's data may generalize to different robot feet as well.

The dataset consists of 43 mostly outdoors images. These are annotated with material class, ground-truth coefficient of friction measured on a humanoid robot foot, and human judgements of friction similar to those in OSA+F. We show the human-sized humanoid robot foot we used in Figure 2. The foot is rigid and its sole is covered with a high-stiffness soft material for shock absorption and an anti-slippage sheet. Locations of the dataset images were chosen such as to cover the same material classes as in OSA+F, as well as extra "dirt", "mud" and "leaves" classes which are common outdoors. At each location, we first measured the maximum friction force by pulling the foot with a spring-scale until it started moving for around 10 trials. We recorded the static coefficient of friction value as the average of the trials divided by the foot's weight. The standard deviation of COF measurements over trials was on average $\sigma = 0.047$. The foot was loaded with a 1.5kg mass and surfaces were checked to be horizontal with a spirit level device. After measuring the coefficient of friction, we removed the foot from the locomotion surface and took pictures of the surface and surroundings using a consumer level camera, along with an annotation of the image location where friction was measured. See Figure 2 for an example picture.

Human judgements of friction were collected as well, using the same procedure as in OSA+F. However, all subjects were given the actual robot foot to look at, feel and experiment on their tables before taking the survey (all subjects' tables were of the same material). The questions were framed as how slippery the subjects expected the surfaces to be in case they were walking on them while wearing the robot's feet as shoes. The subjects responding to this survey were 12 of those who also participated in the OSA+F survey. The dataset contains images of asphalt (3), brick (3), carpet/rug (5), cobble (1), concrete (3), dirt (4), granite/marble (3), leaves (1), linoleum (2), metal (8), mud (1), stone (1), tile (6) and wood (2). We show one example image for each class in Figure 1. Unlike the OSA+F dataset, GTF is not class-balanced. Some material classes are under-sampled, which creates difficulties in training-based algorithms using material class as a feature. In Section IV-D.2 we propose a solution to one of such difficulties: friction prediction without training examples using material class prediction and text mining.

IV. Algorithms for friction from vision

A. Friction from shading

Higher gloss surfaces are usually judged by humans (sometimes mistakenly [5]) as more slippery. Inspired by this observation, we use shading as a feature for friction prediction. Intuitively, we can make an algorithm that analyzes the shading of the scene and classifies a surface as more slippery if it has glossy specular reflections, and less slippery if it is more matte.

We first estimate shading by an intrinsic image decomposition algorithm. These algorithms decompose an original image I into two layers: a shading layer S (irradiance, illumination) and reflectance layer R (albedo, the surface's color). The layers are estimated such that $I = R \cdot S$. Several algorithms exist to estimate this decomposition, such as Retinex [22] or other more complex examples [23]. In this paper we use the Retinex algorithm (implementation in [24]) due to its order of magnitude faster computation time while still achieving high performance [23]. Given an input image, we run Retinex to obtain its shading image and compute the histogram of shading values over the region of interest to estimate friction in that region. We use the the maximum and standard deviation of shading as features:

$$\text{f}_{\text{ShadMax}} = \max_{(i,j)\in C}(S_{i,j}), \tag{1}$$

$$\text{f}_{\text{ShadStd}} = \sqrt{\frac{1}{N}\sum_{(i,j)\in C}(S_{i,j} - \bar{S})^2}, \tag{2}$$

where i, j are indices of the the shading image inside the region of interest C, N is the number of pixels in that region and $\bar{S}$ the region's mean shading. During training, we fit the features to training data using ordinary least squares (OLS) linear regression.

B. Friction from roughness

Humans also use visual estimations of surface roughness to predict friction [6]. Intuitively, frequent variations in image intensity can be used to predict high surface roughness, which is generally associated with high friction.

In this paper we compute the magnitude of the image gradient with a Sobel filter and use the average magnitude of the response as a feature:

$$\mathrm{f}_{\text{GradMu}} = \frac{1}{N} \sum_{(i,j) \in C} \|\nabla I_{i,j}\|, \tag{3}$$

where i, j are indices of the image inside the region of interest C and N the number of pixels in that region. During training, we fit the features to training data using OLS linear regression.

C. Friction from semantic classes

We also use high-level semantic cues for friction estimation based on material, texture and scene classes. We assume these classes are known, given by image-based classifiers such as [15], [16] for material and textures, and [25] for scenes.

Given an input image of material m, we predict friction to be the mean over the training set y on images of the same material:

$$\mathrm{f}_{\text{MatMean}}(m) = \frac{1}{|M|} \sum_{k \in M} y_k, \tag{4}$$

where M is the set of images labeled with material m. When the input image material m is not present on the training set, we use an average friction prior $\mathrm{f}_{\text{MatMean}}(m) = \bar{y}$.

We apply the same logic for texture and scene label features $\mathrm{f}_{\text{TexMean}}$, $\mathrm{f}_{\text{SceMean}}$.

D. Friction from semantic classes and word embeddings

The previous semantic-class method has a disadvantage: it uses a very rough prior for classes not in the training set. For example, if an image-based classifier predicts a surface to be of the material "asphalt" but the friction training set consists only of COF measurements for "concrete" and "ice", the average of the two COF is probably much lower than that of asphalt even though it is intuitively more similar to concrete. We argue that to solve this problem we can use text mining. Text mining methods such as LSA [26] or word embeddings [27], [28] have been used to obtain affordance relations [14] and various other semantic relations [27]. In the case of this paper we are interested in material-material relations such as "asphalt is similar to concrete", and material-slipperiness relations such as "asphalt co-occurs with the word slippery often". We explore both these kinds of relations in this paper through the use of word embeddings.

Word embedding algorithms, such as Word2vec [27] or GloVe [28], embed words into semantic vectors. Each word is represented by a vector of usually 50 to 1000 dimensions, and the cosine similarity between words

$$c_{i,j} = \frac{w_i \cdot w_j}{\|w_i\| \|w_j\|}, \tag{5}$$

is proportional to their co-occurrence in the training set. Here w_i is the vector representing word i. Using the previous example, we can thus estimate the co-occurrence of "asphalt" with "concrete", or even "asphalt" with "slippery" by simple internal products to estimate how similar the two materials are, or how slippery asphalt is. In this paper we trained Word2vec and GloVe models on the complete Wikipedia article dump of 20080103. We chose algorithm parameters by varying them within the ranges recommended in the respective publications, such as to optimize model performance on the semantic tasks described in [28]. Final parameters common to both algorithms were: vector dimension 400 and window size 10. Word2vec-only parameters were: CBOW architecture, negative sampling 10, frequent word sub-sampling 10^{-5}.

After word embeddings are trained, we use semantic similarity queries to estimate friction of an input material. Since the word embeddings exist for all words on the text corpus, we can theoretically estimate friction for thousands of classes. We propose two algorithms for estimating friction using word embeddings.

1) Material-Material similarity: For materials present in the training set this method is the same as the semantic-class method described in Section IV-C. However, when the input image material m is not present on the training set, we use the friction of the "most similar material" $\hat{m}$ in the training set:

$$\mathrm{f}_{\text{WordMM}}(m) = \frac{\bar{y} + \mathrm{f}_{\text{MatMean}}(\hat{m})}{2}. \tag{6}$$

$$\hat{m} = \arg\max_j c_{m,j}, \tag{7}$$

We average $\mathrm{f}_{\text{MatMean}}(\hat{m})$ with the friction prior $\bar{y}$ in order to attenuate errors due to possible wrong material associations.

2) Material-Slipperiness similarity: In this method we estimate friction by word-similarity between the queried material name and a list L of slipperiness-related words[2]. The intuition behind this approach is that the more often a material co-occurs with words such as "slip", "slipped", "slippery" in text then the more likely it is to be slippery for the average contact material. The advantage of the method is that no friction measurements have to be made in order to rank materials by predicted friction, which might be sufficient for some robotic applications (e.g. always plan paths through least slippery options). In this paper we still linearly fit the function to training data, just like with the rest of the features. The feature we propose is the maximum similarity between an input material m and the slipperiness words in list L:

$$\mathrm{f}_{\text{WordMS}}(m) = \max_{j \in L} c_{m,j}. \tag{8}$$

[2]The full list of slipperiness-related words we use is: slipped, slipping, skid, slue, slew, slide, skidded, slued, slided, skidding, slueing, sliding, lubricious, nonstick, slick, slimed, slimy, slithering, slithery. They were obtained by searching and conjugating words related to the word slip on WordNet [29].

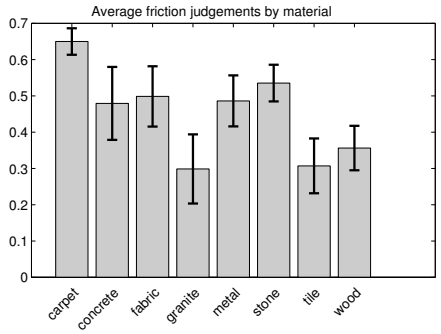

Fig. 3. Average and standard deviation of friction judgements for each material, texture and scene label on the OSA+F dataset.

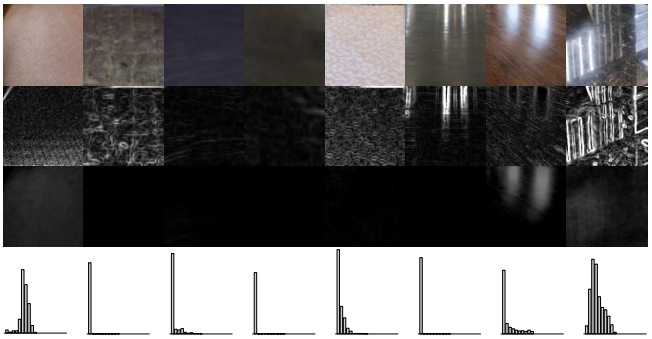

Fig. 4. 8 images from OSA+F sorted from highest to lowest average friction judgements. From top to bottom: original image, gradient image, shading image, histogram of shading image.

V. Results

We now analyze the data collected and friction prediction results. We chose to use two metrics for algorithm evaluation: 1) Root Mean Squared Error (RMSE) between real and predicted friction values on the test set. Results reported are 2-, 5- and 10-fold cross validation values of the RMSE (i.e. average RMSE over the 2, 5 and 10 test sets respectively). All dataset splits are provided together with the datasets. 2) Pearson correlation significance ($p < 0.05$ or $p < 0.01$) between real and predicted friction values on the whole dataset. We use this metric to estimate how chance could be responsible for the correlation between algorithms' predictions and real friction. Due to the relatively small size of the datasets, we choose to report p values on the whole dataset instead of the test sets.

A. OSA+F

Analysis of the dataset

We computed the average and standard deviation of friction judgements for each material, texture and scene. According to a 2-way ANOVA, several relationships between materials are statistically significant: carpet's friction estimates are higher than all other classes; and concrete, fabric, metal and stone are all higher than granite, tile or wood. In the case of textures, the only significant difference is between the labels grid and paisley. Scenes are also poorly informative in this dataset: the only significant difference is between bedroom and foyer. These results indicate material to be a better candidate for prediction of human judgements of friction. We show the average and standard deviation of friction judgements per material in Figure 3.

One recurrent observation in human perception literature is the reliance of humans on gloss to estimate friction. We test this hypothesis on the dataset by a Spearman correlation between friction judgements and gloss/shine estimates as given by the original OpenSurfaces dataset. The Spearman correlation coefficient is $r = -0.344$ ($p < 0.01$), which indicates a significant relationship between the two. However, when computing the correlation independently for each material class, we found that gloss only correlates significantly with friction judgements for the material granite/marble $r = -0.781$ ($p < 0.01$).

Figure 4 shows 8 images of the dataset sorted from highest to lowest mean human friction judgement. For each image we also show data used for image-based features: the gradient image, the shading image and the histogram of values in the shading image. Interestingly, we note that floors with strongly specular reflexions (i.e. higher gloss) are considered the most slippery of the whole dataset, which can be observed in the shading image by larger mean and maximum shading values. The figure also shows that simple single features such as gradient or gloss are insufficient to predict human judgements of friction. For example, the surface with most perceived friction (a carpet) is according to our simple "maximum shading" feature very slippery due to what looks like a glow in the surface.

Algorithm evaluation

We set y, the target function to be predicted, as the average of human friction judgements for each image. In Table I we show the algorithm evaluation results on the OSA+F dataset. We show both 10-, 5- and 2-fold cross validated RMSE values, the algorithms' rank according to the average of the previous three values, and significance of Pearson correlation. The results in this table were obtained assuming ground-truth material, texture and scene classes are known by the algorithms. We use the following two baselines for better comparison and interpretation of the results: 1) "Constant friction" baseline: the mean of y over the training set is used as the prediction; and 2) "Single subject" baseline: we use a single subject's friction estimates as the prediction. We do this for each subject as a predictor and then average the results over all subjects. The objective is to measure performance of a single human in accomplishing the same task as the algorithms (i.e. estimate the average person's friction judgement).

The single subject baseline achieved 0.104 RMSE on 5-fold cross validated results, which was slightly lower than the constant friction baseline (0.137) but indicates high variability among subjects. In fact, inter-subject variability is high (σ =0.166, or 41% of the mean). The best performing algorithms were MatMean and WordMM, which scored 0.083 RMSE. In these experiments MatMean and WordMM are in fact equivalent, since the dataset is class-balanced and

TABLE I

OSA+F: PREDICTING MEAN HUMAN DATA

Features	$RMSE_{CV10}$	$RMSE_{CV5}$	$RMSE_{CV2}$	p	AvgRank
Const	0.137	0.137	0.140		2
SingleSubj	**0.103**	**0.104**	**0.104**	*	1
HumanGloss	0.131	0.129	0.132	*	5
GradMu	0.137	0.138	0.142		8
ShadStd	0.125	0.125	0.129	*	3
ShadMax	0.128	0.128	0.131	*	4
TexMean	0.136	0.137	0.140	*	6
SceMean	0.142	0.134	0.158	*	7
MatMean	**0.081**	**0.083**	**0.086**	*	1
WordMM	0.081	0.083	0.086	*	1
WordMS	0.137	0.138	0.142		9

Note: $p < 0.05$ is marked with *, $p < 0.01$ with **. TexMean, SceMean, MatMean, WordMM and WordMS use ground-truth semantic labels (i.e. of texture, scene, material).

as such all materials are present on the training set. This result matches the previously stated observation that in this dataset material is highly discriminatory.

Interestingly, human judgements of gloss as provided by the original OpenSurfaces dataset scored 0.129 RMSE. The simple statistics of shading images we developed, ShadStd and ShadMax, had a similar but slightly lower error (0.125 and 0.128). All other features either performed at constant baseline level or did not have significant correlation with the mean human judgements. In general, features had similar performance at the different cross validation ratios.

Word embeddings on a larger number of materials

In the previous experiment the method based on material-slipperiness word similarities (WordMS) performed at baseline level and did not correlate significantly with human judgements. The actual correlation of the metric with slipperiness judgements in general is nevertheless difficult to estimate from this dataset due to its small number of materials, which is 8. We conducted one further experiment where we asked 19 new subjects to rank a list of 19 different materials[3] from most to least slippery. The question included only the names of the materials and no supporting images. We computed the average ranking of materials over the subjects and compared this average with the word similarity score given by WordMS (8). The Spearman correlation between the human rankings and f_{WordMS} was a low but significant $r = 0.4607$ ($p < 0.05$). Word embeddings trained on Wikipedia thus seem to encode some knowledge of human judgements of friction, even though at a low correlation level not visible on the OSA+F material classes.

B. GTF

Analysis of the dataset

We did the same analysis as in the OSA+F dataset with GTF data, now targeted at robot locomotion. Figure 5 shows

[3]The complete list was: asphalt, brick, cardboard, carpet/rug, ceramic tile, concrete, fabric/cloth, glass, grass, ice, leather, linoleum, marble/granite, metal, mud, plastic, puddle on asphalt, stone, wood. The subjects were told that with the exception of ice, mud and puddle all materials were dry. Like the original OSA+F task, we also told the subjects to make their judgements assuming they are walking with their normal shoes.

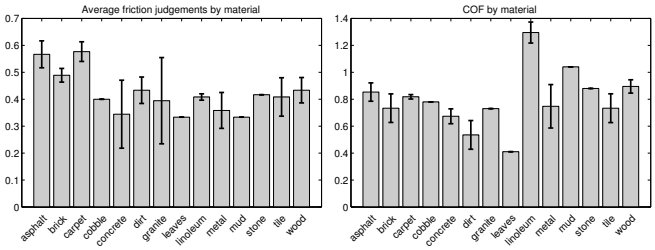

Fig. 5. Average and standard deviation of human judgements of friction (left), and real COF (right) for each material on the GTF dataset.

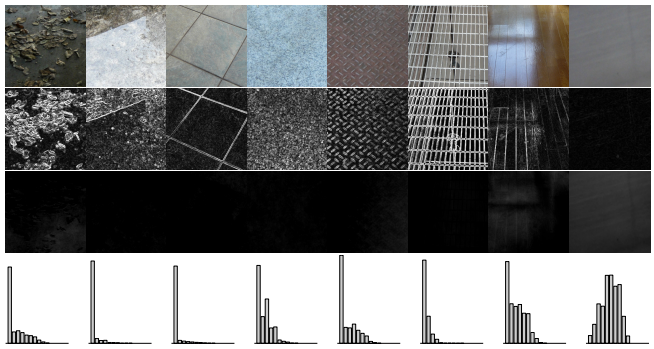

Fig. 6. 8 images from GTF sorted from lowest to highest COF. From top to bottom: original image, gradient image, shading image, histogram of shading image.

the average and standard deviation of friction judgements and real COF for each material. According to a 2-way ANOVA, linoleum had significantly higher friction than all other materials except mud and stone. Wood, asphalt and mud were significantly larger than dirt and leaves. Interestingly, the high COF of mud was not predicted by most human judgements, because contrary to some subjects' intuition mud was sticky rather than slippery. Finally, carpet COF is only significantly higher than dirt.

We also show 8 images of the dataset sorted from lowest to highest COF in Figure 6. We can see how this dataset is more challenging than OSA+F. Surfaces with least friction now include both leaf-covered and wet concrete. The intrinsic image decomposition does not detect specular reflections on the wet case, leaves lead to what could naively look like a surface of high roughness (when in fact leaves can slide easily), glossy wood is actually not slippery for the robot foot because of its anti-slippage sheet, etc. Such examples indicate once again that material classification might be the safest option to friction estimation, although detection of surface "contamination" is crucial as well (e.g. of water, oil, leaves, grain, dust). In fact, one main observation we made during the collection of this dataset was that since the foot is flat, smooth and rigid, its COF is the lowest on contaminated surfaces: clean marble had high friction, dusty was low; small $1mm^2$ stones, leaves, or water drastically reduced the COF.

Algorithm evaluation

For this dataset we set y, the target function to be predicted, as the real COF. In Table II we show the algorithm evaluation results. As in the previous section, the results

showed in this table were obtained assuming ground-truth material is known. We use the following two baselines for better comparison and interpretation of the results: 1) "Constant friction", and 2) "Single subject" baselines are the same used for the evaluation of OSA+F. "Single subject" estimates are thus human judgements of friction, while the target function y is the real COF. The motivation for evaluating this metric is to find out whether an average inexperienced robot operator, even if familiarized with the robot's foot, can predict or not the friction coefficient between the robot and ground. This has of course strong implications for the design of control architectures and interfaces for remotely controlled robots. Finally, we also evaluate another baseline: 3) "Mean Subject" the average of the subjects' friction judgements. Thus, we measure how much a group of inexperienced robot operators, instead of a single operator, can help predict COF. The motivation is to compare this metric with the single subject metric, thus helping to understand whether an increase in robot operators (e.g. crowd-sourced operators) may increase prediction performance.

Constant friction baseline in this dataset achieves 0.194 RMSE on 5-fold cross validate results. Perhaps surprisingly, single subject judgements of friction achieve performance roughly equal to constant baseline, meaning they are poorly predictive of real COF in this dataset. Also, using multiple subjects (MeanSubj) did not improve performance considerably when compared to the average result obtained with a single subject. Image features (GradMu, ShadStd, ShadMax) were roughly as predictive as human judgements, actually up to 6 % better. However, the image features' correlation with real COF was only significant for ShadMax, which was also a good predictor in the human data of the OSA+F dataset.

Once again material classification, MatMean, was the highest scoring method, achieving 0.137 RMSE on 5-fold cross validation. WordMS further improves performance by around 2% since it deals with classes unseen on the training set. Interestingly, our material-slipperiness word similarity method WordMS achieved higher (and statistically significant) performance when compared to both human judgements and image features. Results shown in Table II for WordMM and WordMS were obtained using the Word2vec algorithm for word vector training. We also evaluated performance on a different word embedding algorithm, GloVe [28], which is together with Word2vec currently one of the best performing on semantic tasks [30]. On average, the RMSE on GloVe-trained vectors was 3% higher.

TABLE II
GTF: PREDICTING COF

Features	$RMSE_{CV10}$	$RMSE_{CV5}$	$RMSE_{CV2}$	p	AvgRank
Const	0.188	0.194	**0.182**		3
SingleSubj	0.176	0.187	0.189		2
MeanSubj	**0.174**	**0.186**	0.187		1
GradMu	0.172	0.180	0.176		5
ShadStd	0.177	0.191	0.196		6
ShadMax	0.171	0.187	0.182	**	4
MatMean	0.130	0.137	**0.134**	*	1
WordMM	**0.127**	**0.135**	0.141	*	2
WordMS	0.155	0.170	0.180	**	3

Note: $p < 0.05$ is marked with *, $p < 0.01$ with **.

VI. CONCLUSION AND DISCUSSION

Friction prediction from visual cues is a challenging but crucial problem for robot locomotion. In this paper we described and analyzed two new open datasets for friction estimation. We collected both COF and human judgement data in order to provide room to inform both robotics and human perception communities. Lastly we proposed and evaluated a set of methods for friction prediction. Overall, we importantly observe:

Robot teleoperation. Human judgements have low predictive power of COF in the GTF dataset, meaning it might be a wrong choice to trust slipperiness judgement to inexperienced robot operators even if they are familiarized with the robot's foot. We can also imagine a robot operation setup where several perception decisions are crowd-sourced over a group of operators. However, even using the mean of 12 subjects as a predictor leads to lower performance than image-based statistics. Constant-friction baselines might actually be safer than human guesses according to 2-fold cross validated results. The observation matches recent findings in the human literature [6] where COF was difficult to estimate for humans. Our proposed image-based feature related to gloss, the maximum image shading, obtained better, significantly correlated, performance than humans. While friction prediction based on material class was the best performing method, the material classification task is still challenging for state-of-the-art computer vision algorithms (70% accuracy [15], [16]). Thus, one way a robot teleoperator could assist the procedure could actually be by material labeling.

COF prediction errors. Material was the most predictive feature for both COF (0.130 RMSE) and human judgements of friction. Image features based on intrinsic shading images performed worse (0.171 RMSE) but slightly better than baseline. Both in this paper and others relying on material classification for predicting friction (e.g. [13]), problems may arise when new materials are traversed. Thus, we proposed methods based on text mining for friction estimation of previously unseen material classes. Matching new materials to trained ones by material-material similarity improved performance by 2%. Estimating friction of a material by the co-occurrence of the material with slipperiness-related words in text was better (0.155 RMSE) than image-based statistics and human subjects at COF-prediction.

Text mining and word embeddings. Algorithms based on text mining may compensate for lack of robot experience in novel scenarios, and are also likely to improve their performance as Natural Language Processing algorithms improve. An interesting open problem is to find ways to adapt the methods based on text mining we proposed here. One important improvement would be to estimate friction between two specific materials. As proposed here, WordMS estimates friction from co-occurrences between material and

slipperiness-related words. Therefore, it obtains not an estimate of friction between two specific materials, but an average estimate of friction of the reference material with all materials which co-occur with it in text.

Human perception. We also replicated recent results in the human perception literature, correlating human judgements of friction and surface gloss/shine [5]. However, we found that this correlation was only significant for the marble material (but not, for example, for tiles). We hypothesize that humans rely on illumination-based features only for certain materials for which it might be predictive.

Research directions. Interesting problems to further explore on the friction-from-vision topic and in the OSA+F/GTF datasets include new prediction methods; further testing of hypothesis regarding human perception of friction; using surrounding context (e.g. objects in the scene); and building new very large crowd-sourced datasets. While building new, larger, completely robot-acquired datasets would be advantageous for the field and allow the application of methods based on deep neural networks [16], [15], several challenges still lie ahead since autonomous locomotion in varied terrain by complex robots is still an open problem.

REFERENCES

[1] K. Kaneko, F. Kanehiro, S. Kajita, M. Morisawa, K. Fujiwara, K. Harada, and H. Hirukawa, "Slip observer for walking on a low friction floor," in *Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference on*, Aug 2005, pp. 634–640.

[2] N. Okita and H. Sommer, "A novel foot slip detection algorithm using unscented kalman filter innovation," in *American Control Conference (ACC), 2012*, June 2012, pp. 5163–5168.

[3] M. Bloesch, C. Gehring, P. Fankhauser, M. Hutter, M. Hoepflinger, and R. Siegwart, "State estimation for legged robots on unstable and slippery terrain," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, Nov 2013, pp. 6058–6064.

[4] J. H. Park and O. Kwon, "Reflex control of biped robot locomotion on a slippery surface," in *2001 IEEE International Conference on Robotics and Automation*, vol. 4, 2001, pp. 4134–4139 vol.4.

[5] A. S. Joh, K. E. Adolph, M. R. Campbell, and M. A. Eppler, "Why walkers slip: Shine is not a reliable cue for slippery ground," *Perception & Psychophysics*, vol. 68, no. 3, pp. 339–352, 2006.

[6] M. F. Lesch, W.-R. Chang, and C.-C. Chang, "Visually based perceptions of slipperiness: Underlying cues, consistency and relationship to coefficient of friction," *Ergonomics*, vol. 51, no. 12, pp. 1973–1983, 2008, pMID: 19034787.

[7] K. W. Li, W.-R. Chang, T. B. Leamon, and C. J. Chen, "Floor slipperiness measurement: friction coefficient, roughness of floors, and subjective perception under spillage conditions," *Safety Science*, vol. 42, no. 6, pp. 547 – 565, 2004.

[8] H. Liu, X. Song, T. Nanayakkara, K. Althoefer, and L. Seneviratne, "Friction estimation based object surface classification for intelligent manipulation," in *IEEE ICRA 2011 workshop on autonomous grasping, Shanghai*, 2011.

[9] M. Tremblay and M. Cutkosky, "Estimating friction using incipient slip sensing during a manipulation task," in *Robotics and Automation, 1993. Proceedings., 1993 IEEE International Conference on*, May 1993, pp. 429–434 vol.1.

[10] K. Hang, F. Pokorny, and D. Kragic, "Friction coefficients and grasp synthesis," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, Nov 2013, pp. 3520–3526.

[11] S. Kajita, K. Kaneko, K. Harada, F. Kanehiro, K. Fujiwara, and H. Hirukawa, "Biped walking on a low friction floor," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 4, Sept 2004, pp. 3546–3552 vol.4.

[12] M. Brandao, K. Hashimoto, J. Santos-Victor, and A. Takanishi, "Footstep planning for slippery and slanted terrain using human-inspired models," *IEEE Transactions on Robotics*, vol. 32, no. 4, pp. 868–879, Aug 2016.

[13] A. Angelova, L. Matthies, D. Helmick, and P. Perona, "Slip prediction using visual information," in *Proceedings of Robotics: Science and Systems*, Philadelphia, USA, August 2006.

[14] Y.-W. Chao, Z. Wang, R. Mihalcea, and J. Deng, "Mining semantic affordances of visual object categories," in *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, June 2015, pp. 4259–4267.

[15] S. Bell, P. Upchurch, N. Snavely, and K. Bala, "Material recognition in the wild with the materials in context database," *Computer Vision and Pattern Recognition (CVPR)*, 2015.

[16] M. Cimpoi, S. Maji, and A. Vedaldi, "Deep filter banks for texture recognition and segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, 2015, pp. 3828–3836.

[17] V. Vineet, O. Miksik, M. Lidegaard, M. Nießner, S. Golodetz, V. A. Prisacariu, O. Kähler, D. W. Murray, S. Izadi, P. Perez, and P. H. S. Torr, "Incremental dense semantic stereo fusion for large-scale semantic scene reconstruction," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2015.

[18] G. Cappellini, Y. P. Ivanenko, N. Dominici, R. E. Poppele, and F. Lacquaniti, "Motor patterns during walking on a slippery walkway," *Journal of Neurophysiology*, vol. 103, no. 2, pp. 746–760, 2010.

[19] S. Bell, P. Upchurch, N. Snavely, and K. Bala, "OpenSurfaces: A richly annotated catalog of surface appearance," *ACM Trans. on Graphics (SIGGRAPH)*, vol. 32, no. 4, 2013.

[20] LimeSurvey Project Team, Carsten Schmitz, *LimeSurvey: An Open Source survey tool*, LimeSurvey Project, Hamburg, Germany, 2012. [Online]. Available: http://www.limesurvey.org

[21] R. W. Fleming, C. Wiebel, and K. Gegenfurtner, "Perceptual qualities and material classes," *Journal of Vision*, vol. 13, no. 8, p. 9, 2013.

[22] E. H. Land and J. J. McCann, "Lightness and retinex theory," *J. Opt. Soc. Am.*, vol. 61, no. 1, pp. 1–11, Jan 1971.

[23] S. Bell, K. Bala, and N. Snavely, "Intrinsic images in the wild," *ACM Trans. on Graphics (SIGGRAPH)*, vol. 33, no. 4, 2014.

[24] N. Limare, A. B. Petro, C. Sbert, and J.-M. Morel, "Retinex Poisson Equation: a Model for Color Perception," *Image Processing On Line*, vol. 1, 2011.

[25] B. Zhou, A. Lapedriza, J. Xiao, A. Torralba, and A. Oliva, "Learning deep features for scene recognition using places database," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 487–495.

[26] T. K. Landauer and S. T. Dumais, "A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge." *Psychological review*, vol. 104, no. 2, p. 211, 1997.

[27] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems 26*, C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, Eds. Curran Associates, Inc., 2013, pp. 3111–3119.

[28] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543.

[29] "Princeton University "About WordNet." WordNet. Princeton University," 2010. [Online]. Available: http://wordnet.princeton.edu

[30] T. Schnabel, I. Labutov, D. Mimno, and T. Joachims, "Evaluation methods for unsupervised word embeddings," in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2015, pp. 0–0.