

LaserShoes: Low-Cost Ground Surface Detection Using Laser Speckle Imaging

Zihan Yan*

Zhejiang University, China
MIT Media Lab, United States
yizhan@media.mit.edu

Yuxiaotong Lin*

Zhejiang University, China
linyxt@zju.edu.cn

Guanyun Wang†

Zhejiang University, China
guanyun@zju.edu.cn

Yu Cai
Zhejiang University, China
cai_yu@zju.edu.cn

Peng Cao
MIT CSAIL, United States
pengcao@mit.edu

Haipeng Mi
Tsinghua University, China
mhp@tsinghua.edu.cn

Yang Zhang
University of California, Los Angeles,
United States
yangzhang@ucla.edu

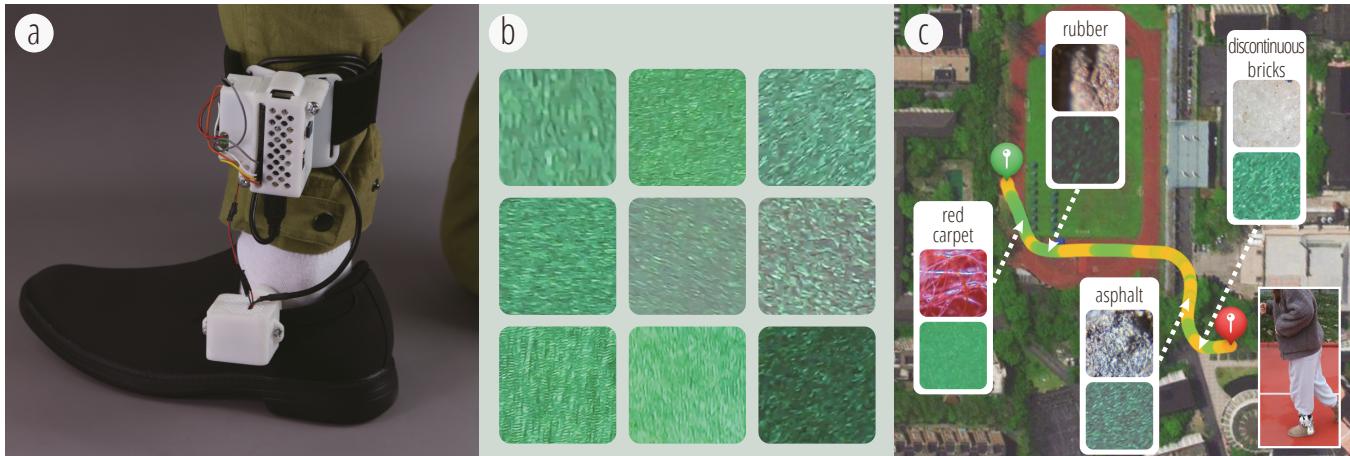


Figure 1: LaserShoes is a ground surface detection system based on wearable laser speckle imaging. (a) The hardware of *LaserShoes*, which consists of two major components: 1) a detecting component, which consists of a laser emitter and an image sensor and is set on shoes; and 2) a processing and assistant component, which mainly consists of a Raspberry Pi. The processing and assistant component is attached to a user's lower leg; (b) To detect ground surfaces, *LaserShoes* recognizes patterns exhibited by laser speckles induced on different ground surfaces. (c) One example application of *LaserShoes* is personal running assistant. *LaserShoes* can identify ground surfaces on which the user is running and log this information for running analysis.

ABSTRACT

Ground surfaces are often carefully designed and engineered with various textures to fit the functionalities of human environments

*The first and second authors contributed equally to the paper.

†Corresponding author: Guanyun Wang, guanyun@zju.edu.cn.



This work is licensed under a Creative Commons Attribution International 4.0 License.

and thus could contain rich context information for smart wearables. Ground surface detection could power a wide array of applications including activity recognition, mobile health, and context-aware computing, and potentially provide an additional channel of information for many existing kinesiology approaches such as gait analysis. To facilitate the detection of ground surfaces, we present *LaserShoes*, a texture-sensing-enabled system using laser speckle imaging that can be retrofitted to shoes. Our system captures videos of speckle patterns induced on ground surfaces and uses pre-processing to identify ideal images with clear speckle patterns collected when users' feet are in contact with ground surfaces. We demonstrated our technique with a ResNet-18 model and achieved real-time inference. We conducted an evaluation in different conditions and demonstrated results that verified the feasibility.

CCS CONCEPTS

- Human-centered computing → Ubiquitous and mobile devices.

KEYWORDS

Smart Shoes; Laser Speckle Imaging; Surface Identification; Texture Identification; Context Aware Computing

ACM Reference Format:

Zihan Yan, Yuxiaotong Lin, Guanyun Wang, Yu Cai, Peng Cao, Haipeng Mi, and Yang Zhang. 2023. LaserShoes: Low-Cost Ground Surface Detection Using Laser Speckle Imaging. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23), April 23–28, 2023, Hamburg, Germany*. ACM, New York, NY, USA, 20 pages. <https://doi.org/10.1145/3544548.3581344>

1 INTRODUCTION

Human environments contain rich contextual information that could be used to power a variety of context-aware computing applications. Users' presence in a kitchen, for example, often indicates food preparation activities, whereas classrooms indicate learning and theaters indicate entertainment. As a result, accurate and robust sensing of user presence in environments with varying functionalities has long been desired in HCI [1, 70, 71]. Additionally, fine-grained information on user location could also facilitate conventional sensor-aided approaches such as gait analysis [14, 30], activity logging [26], and beyond for medical research and many more in-the-wild studies.

In this research, we create a wearable system to recognize ground surfaces, which are a universal and expressive feature of human environments and often are strong indicators of user contexts. Surface texture, as a distinguishing feature of any ground surface defined by four characteristics including lay, flaw, roughness, and waviness [43], has recently received tons of attention in the sensing research field. For example, texture-based ground surface detection has been widely used in applications of robotics, such as assisting mobile robots in detecting obstacles [73] and promoting autonomous agriculture [42].

As we lay barefoot on ground surfaces, we feel the soft grass of a lawn, lumpy fabrics on a carpet, gritty soil of a hiking trail, smooth tiles of a bathroom, grainy wood of a floor, and rough sands on a beach. We believe that wearable intelligence could benefit from enhanced perceptual capabilities of sensing ground surfaces, similar to what humans can do but without limitations in sensitivity, granularity, latency, and time of operation, in order to achieve a better understanding of environments and user contexts, and to provide assistance, accommodate for natural interactions, and log important patterns in information for analysis and diagnosis.

As users' feet are almost always in contact with ground surfaces, shoe-instrumented wearables serve as an ideal platform for sensing ground surfaces. To enable shoe wearables to sense ground surfaces, we propose *LaserShoes*, a low-cost ground surface detection system using the laser speckle imaging technique (Fig. 1). In comparison with conventional vision-based approaches taking RGB photos of ground surfaces, laser speckle imaging reveals richer and more accurate information about textures of ground surfaces using an active signal – laser beams. When compared to cameras, laser speckle imaging can distinguish surface textures that appear

visually similar. Additionally, unlike conventional imaging systems which require lenses, laser speckle imaging does not require a lens and thus cannot provide clear visuals of users' backgrounds to preserve privacy.

Our system mainly consists of a laser emitter, an image sensor (CCD), and a Raspberry Pi board. The laser emitter and the image sensor are connected to shoes to capture videos of speckle patterns that reflect surface textures. The Raspberry Pi board is instrumented to a user's lower leg and runs the detection pipeline which features a pre-processing phase to eliminate blurry images, and a deep learning model to acquire ground surface types. The entire system costs \$136. We recruited 15 participants in a user study where they were asked to walk on 24 ground surfaces for 1~2 minutes. In total, we collected 28,492 1.5s video sessions. We validated our system under within-user and cross-user conditions, and the classification accuracy of within-user and cross-user conditions is 86.93% and 80.57%, respectively. We also carried out three additional studies to tease out the performance of our system in detecting dry, wet, and frozen surfaces, and sand surfaces of different grain sizes, and under various lighting conditions. Finally, we demonstrated applications enabled by our system, such as personal running assistant, gait analysis, surface-aware cleaning equipment, coarse navigation, and daily activity recognition through localization.

In summary, our main contributions include:

- We designed and implemented *LaserShoes*, a laser-imaging-based ground surface detection wearable system that can identify ground surfaces.
- We designed a data process method for *LaserShoes* to identify relative stationary frames from collected videos and completed an end-to-end real-time inference pipeline based on contemporary deep learning techniques.
- We conducted an evaluation with 15 participants to investigate the performance of *LaserShoes* with two validation mechanisms (i.e., within- and cross-user), and under various surface and environmental conditions.

2 RELATED WORK

2.1 Sensing Ground Surface with Smart Shoes

With the rise of ubiquitous computing, various smart shoes are designed and developed for sensing ground surfaces and accommodating novel interaction modalities [62, 67]. Taking advantage of their unique position, and linking the foot with ground surfaces, smart shoes can often yield information beyond what is possible through wearables at other body locations.

Prior work has demonstrated ground surface identification using foot kinematics, which could be used for danger alerts and human activity recognition. Specifically, Otis et al. [37] used a variety of sensors, including accelerometers, gyroscopes, and force sensors, to distinguish between the physical properties of different soils. Cheng et al. [8] designed wearable capacitive sensors and applied them around users' ankles to recognize whether users were walking on concrete or in the meadow. Furthermore, Matthies et al. [31] designed CapSoles leveraging the capacitive ground coupling effect to detect six different ground surfaces. Zrenner et al. [75] revealed the relationship between foot kinematics and ground surfaces with different properties using inertial measurement units (IMU). Strada

et al. [50] used gait data collected by inertial sensors embedded in the shoes' soles to identify surface types and conducted experiments on four different ground surfaces. However, foot kinematics can be largely affected by the user's intrinsic walking characteristics [31] and health conditions [41]. In contrast, *LaserShoes* uses laser speckle imaging to recognize ground surfaces of different textures, which is immune to differences in gait, and has been evaluated on 24 different ground surfaces and demonstrated robustness under a variety of environmental conditions.

2.2 Surface Texture Detection Techniques

Surface texture is a complex condition resulting from a combination of roughness (nano and micro-roughness), waviness (macro-roughness), and lay and flaw [43]. Surface texture recognition has been widely used in various application domains [2, 6, 35, 49]. Conventional approaches to texture recognition include microscopes and roughness meters which are expensive and often stationary, making them difficult to be instrumented on a user's body as wearable devices. It is also possible to use tactile sensors with a portable form factor to identify surface textures [12, 38, 57, 60, 66]. These tactile sensors can augment the sensation of touch and assist surface texture recognition.

Closer to our system, several prior works have investigated optical approaches for surface texture recognition. For example, Su et al. [53] acquired subsurface scattering characteristics measured by time-of-flight (ToF) cameras to identify surface texture features. Researchers also combined a multi-spectral light source and an image sensor to recognize surface textures [20, 65]. These techniques, however, often rely on complex devices or multiple light sources, which are expensive to scale. In this research, we chose laser speckle imaging, a relatively low-cost approach consisting of mainly a laser and an image sensor, to capture surface texture features at high fidelity. We furthered this sensing approach into an end-to-end system and evaluated it with realistic surface and environmental conditions.

2.3 Laser Speckle Imaging

LaserShoes is closely related to prior works on Laser speckle imaging [19], which is a technique that uses an image sensor to obtain patterns in laser speckle images corresponding to surface textures when a beam of coherent light, such as a laser, illuminates the surface. This method has been used in a variety of fields. In the medical field, for example, it is used to monitor capillary perfusion in human skin tissues and brain blood flow maps in rodents [10, 11, 16]. In HCI, Laser Speckle Imaging has been used to recognize appliance use and home activities [54, 72] to achieve motion sensing and motion tracking [36, 48, 74]. And a non-contact force sensing can also be achieved by applying Laser Speckle Imaging to manifest surface deformation which is corresponded to force [40]. Furthermore, Laser Speckle Imaging techniques can be used to expose surface characteristics for surface material identification. SpecTrans [45] leverages Laser Speckle Imaging in conjunction with multi-spectral LED illumination to classify textureless, specular, and transparent materials for interactivity. Laser Speckle Imaging is highly sensitive and can even reveal small composition differences of materials that appear identical to human eyes. SensiCut [13], for example, applies

this technique on a laser cutting machine to identify the pending materials before cutting, to improve its safety and workflow.

Our work leveraged this sensing approach to identify ground surfaces, a drastically different class of surfaces than the ones in the prior work. Our different application scenario comes with unique challenges such as the relative motion between a user's feet and the ground surface. To overcome these challenges, we developed an end-to-end wearable system with a custom pre-processing phase to filter out blurry speckle images due to the motion effect, resulting in a robust system that we evaluated with a wide range of common ground surface types.

3 PRINCIPLES OF OPERATION

LaserShoes is based on two principles of operation: 1) we used Laser Speckle Imaging to detect ground surface textures, and 2) we used the variance of grayscale-converted frames from recorded videos to infer gait status and obtain speckle images with high quality.

First, Laser Speckle Imaging can reveal surface texture characteristics. When a beam of coherent light (e.g., laser) illuminates a ground surface, the light will be reflected, and captured by a nearby image sensor, forming an image with laser speckles, as shown in Fig. 2 (a). This phenomenon occurs because ground surfaces are rough – the micro geometry of ground surfaces varies the optical paths of the laser beam. Thus, each pixel of the image sensor will receive the reflected laser beam with different constructive and destructive interference, forming laser speckles. Because different ground surfaces have different micro geometries, the resulting laser speckle patterns vary and could be leveraged to identify ground surfaces.

Second, we applied Laser Speckle Imaging with the consideration that a user's feet could be in constant motion (e.g., walking and running) in relation to ground surfaces. The sensor's movements relative to the ground manifest as the motion effect on images, resulting in blurry laser speckle images that have lower variances compared with those that have sharp speckles. As illustrated in Fig. 2 (b), the laser speckle images are much clearer when a user's foot is in contact with the ground than when the foot is moving in the air. We utilized the variances of grayscale speckle images to identify the foot-ground contact period from recorded videos and used only speckle images collected from this period for the subsequent classification.

4 HARDWARE DESIGN

We prototype *LaserShoes* to investigate the capabilities of laser imaging in ground surface detection. Although our current implementation is relatively bulky and impractical for direct adoption, our end-to-end prototype enables us to effectively verify our sensing principle, conduct technical evaluation, and explore potential applications. The form factor of our current prototype is akin to established works in the HCI community [7, 61, 68]. In this section, we introduce our hardware configurations and fabrication.

4.1 Embedded System

We apply Laser Speckle Imaging to capture speckle patterns and recognize ground surfaces. The technique has been used in the HCI community and could be eye-safe [4]. To utilize this technique, our system consists of four parts: 1) a laser emitter, 2) an image

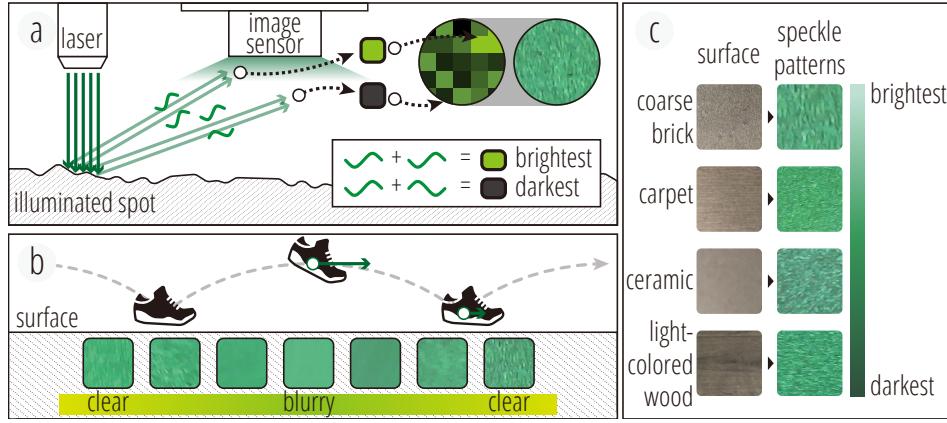


Figure 2: Two principles of operation and speckle patterns induced by different ground surfaces. (a) The principle of Laser Speckle Imaging. The optical paths of laser beams vary due to variances of the surface micro geometry, resulting in constructive and destructive interference on a nearby image sensor; (b) The gait status affects the blurriness of the laser speckle images. When a user's foot moves in the air, the corresponding laser speckle images are blurry, whereas when the foot comes into contact with the ground surface the corresponding laser speckle images are clear; (c) Ground surfaces and the induced laser speckles. The speckle images measure 256×256 pixels.

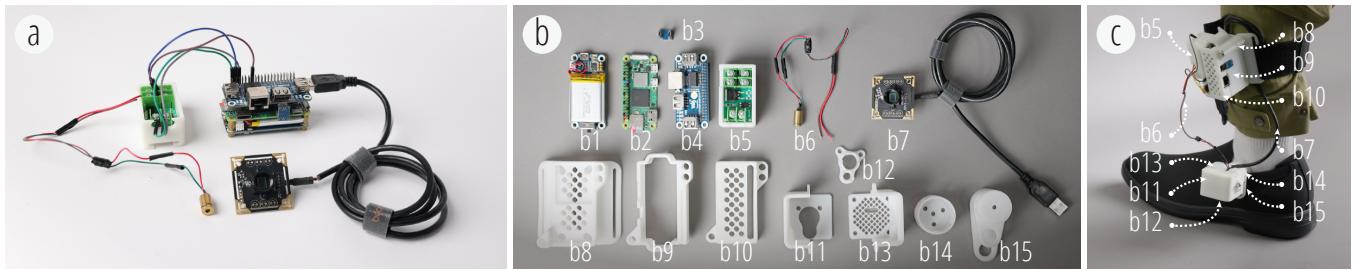


Figure 3: The hardware of *LaserShoes*. (a) The circuit connection of electronic components; (b) Individual unit of the system, in which b1-b7 are electronic components, b8-10 are mechanical components for housing the Raspberry Pi and other assistant modules, and b11-15 are the mechanical components for housing the laser emitter and the image sensor and affixing our system to shoes. b1-battery module, b2-Raspberry Pi Zero 2 W, b3-connector between b2 and b4, b4-USB interface module, b5-switch module for the laser emitter power supply, b6-laser emitter, b7-image sensor, b8-support component between a user's lower leg and the hardware, b9-square housing to cover the Raspberry Pi and assistant modules, b10-top lid of b9, b11-one of the semi-cubic shell of the container, b12-fixture for the laser emitter, b13-another semi-cubic shells of the container, b14-cylindrical housing that connect the container and the clamping part, b15-clamping part that attaches this structure set to shoes with a series of holes for adjusting the angle of the container to it using b14; (c) *LaserShoes* worn on a user's foot and lower leg with all components annotated.

sensor, 3) a Raspberry Pi board, and 4) assistant modules. The laser emitter and the image sensor compose the detecting component, while other parts compose the processing and assistant component. The hardware details of our system are shown in Fig. 3. Compared to prior works [20, 65], the core sensors bundled in our system are more compact to set on shoes. The enclosure of the system is 3D printed using photosensitive resin. The entire system and its manufacturing cost are \$135.23, and the combined cost of the laser emitter and image sensor is \$23.14. The cost of each component is shown in Table 1.

Laser Emitter. We select a laser emitter with a 520nm wavelength and 5mW output power based on our configuration experiments (see Section 4.2.2). Given that using a low-power laser emitter will

result in insufficient illumination and unclear speckle patterns, and that using a high-power laser may not be eye-safe, we ultimately choose a 5mW laser (Class IIIA) which is chronic viewing hazard but safe for transient exposures. Additionally, in order to have maximum laser reflection to preserve signal-to-noise ratio (SNR), we set the laser emitter vertical to ground surfaces.

Image Sensor. Given that our system is mounted on users' shoes, it is subject to movement as users walk, leading to the loss of speckle information in parts of the image due to motion blur. To extract images with clear speckle patterns from captured videos, we select an OV2710 image sensor with a relatively high frame rate of 60 fps. We set the resolution of the image sensor as 1280×720 pixels,

Table 1: Costs of main components of *LaserShoes*.

| Module | Laser Emitter | Image Sensor | Switch Module | Raspberry Pi Board with USB interface board | Battery Module | Fabrication |
|------------|---------------|--------------|---------------|---|----------------|-------------|
| Price (\$) | 7.00 | 16.14 | 0.52 | 76.86 | 17.57 | 17.14 |

which is the highest resolution under the 60-fps frame rate. It is worth noting that our system does not use a lens because laser beams reflected by ground surfaces are always in focus, resulting in sharp speckle patterns that are distributed uniformly across the captured images when a user's shoe is relatively still with respect to ground surfaces. To further improve SNR, the image sensor is placed right next to the laser emitter.

Raspberry Pi Board and Assistant Modules. For image acquisition and processing, we choose the Raspberry Pi Zero 2 W, for its compact size, superior speed, and wireless connectivity. With the connected laser emitter and image sensor, the Raspberry Pi board carries out three functions: 1) supplying power to the laser emitter from GPIO, 2) acquiring videos from the image sensor through a USB interface and 3) processing acquired videos and yielding the detected type of ground surface to users. The assistant modules include a battery module, a USB interface module, and a switch module to safely supply power to the entire system.

4.2 Configurations

In order to identify the optimal configuration of our system, we conducted experiments using various combinations of *laser wavelengths* and *distances*, as they are two significant factors affecting the formation of laser speckles, and investigated their performance in surface classification. In these experiments, we used an image sensor which was a model commonly used on webcams with a pixel size of $3\mu\text{m} \times 3\mu\text{m}$.

4.2.1 Image sensor. Given that our system operates in a moving scenario, an image sensor with a sufficient frame rate is required to ensure the quality of captured videos and to extract clear speckle patterns from those videos. Through experiments in which we collected videos while researchers with the camera configured at different frame rates were walking at their normal speed, we discovered that the standard 30-fps frame rate is insufficient due to the motion effect, resulting in an excessive number of blurry images. On the other end, sensors with higher frame rates are often costly, which contradicts our design goal of being low-cost. As a result, we choose a frame rate of 60 and rely on a custom pre-processing pipeline to mitigate the motion blur (see in Section 5.1).

4.2.2 Wavelength and distance. Since infrared lasers are difficult to debug, we selected wavelengths of laser in the visible spectrum. Specially, in our experiments, we investigated 4 different representative laser wavelengths (405nm, 450nm, 520nm, and 650nm). In terms of distance, considering that our system is intended to be fixed on shoes, which often hold a relatively short distance with ground surfaces, we kept the distance as short as possible while maintaining sufficient clearance for the light path (i.e., from the emitter to ground surfaces and back to the image sensor). Thus, in this case, for each wavelength, we investigated its performance

at distances with the ground surfaces of 1cm, 3cm, 5cm, 7cm, 9cm, 11cm, 13cm and 15cm (Fig. 4).

For each wavelength-and-distance combination, we collected a number of images with speckle patterns on five surfaces (wood, fabric, concrete, rubber, and ceramic). During the collection, we manually swapped the laser emitter of different wavelengths and adjusted the sensor distance to the ground surface. In order to evaluate the qualities of these images, we conducted a quick validation using ResNet-18 [21], with collected images split into a training set and a testing set. Our assumption is that laser speckle images with high-quality speckle patterns will yield relatively high classification accuracy, revealing optimal wavelength-and-distance combinations.

The average classification accuracies and their standard deviations of all wavelength-and-distance combinations are shown in Appendix A. Results indicate that the green laser (520nm) exhibits both high accuracy and stability, though almost all combinations reach high classification accuracies. When the distance is under 11cm, the accuracies of the green laser are all above 98%. Thus, in our subsequent studies, we choose the green laser with a 520nm wavelength and set the distance between the sensor and ground surfaces to under 11cm when affixing the sensor to users' shoes.

4.3 Mechanical Structure and Fabrication

We build a mechanical structure of two modules that can achieve angle adjustment of the detecting component to ground surfaces and the fixation of the system on a user's leg (Fig. 3). The first module consists of five parts: two semi-cubic shells forming a container (b11, b13), a limiter with two cylindrical channels (b12), a cylindrical housing (b14), and a clamping part (b15). The two semi-cubic shell surfaces are joined together into a cube container by screws on the side. The image sensor is fixed inside the cube housing via slots in the four corners of the cube container's inner side, and the laser is fixed on the bottom side of the cube housing via a fixture (b12). A number of rivet structures are used to connect the cube container to the column housing b14, and to implement the rotatable connection between the column housing b14 and the clamping part (b15). Screws are used to secure a series of discontinuous holes in the column housing and the clamping part, allowing an adjustable angle between the cube container and the clamping part, ranging from 0 to 90 degrees in a 15-degree step. As the clamping part of the first module is fixed to the outer side of a user's ankle, adjusting the angle between the cube container and the clamping part changes the angle between the laser sensing beam with the user's leg and thus with the ground surfaces.

The second module contains four parts: a supporting part (b8), a square housing (b9), a top lid (b10), and a controller box (b5). Among these, b8, b9, and b10 are jointed by three studs on the corners to form a container for the combined structure of the Raspberry Pi board and the battery module. The container measures approximately 65.7mm in length, 30.6mm in width, and 46.0mm in height.

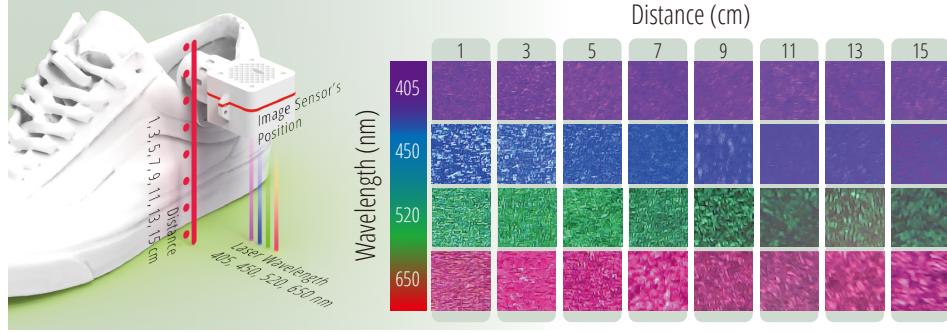


Figure 4: 32 wavelength-and-distance combinations that we selected and the corresponding sample speckle imaging data of each wavelength-and-distance combination. We selected four different wavelengths of 405nm, 450nm, 520nm and 650nm and eight different distances of 1cm, 3cm, 5cm, 7cm, 9cm, 11cm, 13cm and 15cm. When collecting data, we adjusted the wavelength by swapping laser emitters manually.

The USB port and the charging port are reserved for the exterior of the container. The controller box (b5) contains the switch module and is attached to the rest of the module with a side slide. This module is fixed to the outside of the user's lower leg with straps fitting through b8 and the main structure of the container is kept away from the user's skin to avoid possible discomfort due to the heat dissipation of our system. The above mechanical structures are 3D printed with photosensitive resin at a 0.05 mm resolution using a Lite600HD 3D printer.

5 GROUND SURFACE DETECTION

The whole ground surface detection pipeline of *LaserShoes* is illustrated in Fig. 5. *LaserShoes* device is expected to work despite the constant motion with ground surfaces while users are walking. Every 90 frames are treated as a video session, taking about 1.5 seconds to collect. This duration is selected for our observation that at least one foot-ground contact would appear in the video session when users walk at normal speeds.

Video sessions are fed into our ground surface detection system, which consists of a pre-processing phase and a deep learning model for classification. Specifically, with this pre-processing phase, we select images with clear speckle patterns from the collected videos and crop selected speckle images into small images before feeding them into a deep learning model for classification, as a data enhancement technique to increase our data collection efficiency. This pre-processing phase allows *LaserShoes* to deal with distance change and motion blur caused by users' gait.

5.1 Data Pre-processing

The motion of users' feet causes the speckle patterns to be blurry and thus contain little information on ground surfaces (Fig. 2 (b)). To achieve high detection accuracy, it is necessary to extract high-quality images with clear speckle patterns. Our pre-processing phase contains four stages (Fig. 5 (b)-(e)), including 1) identifying the foot-ground contact periods, 2) cropping images, 3) removing partial blurry images, and 4) removing fuzzy patterns. Specifically, we first identify images collected from foot-ground contact periods. We then crop these foot-ground contact images into small images with the size of 256 × 256. We discard cropped images with partial

blur or fuzzy patterns. After the pre-processing phase, we obtain a group of cropped images with clear speckle patterns to feed into our deep-learning model. The details of each stage of this pre-processing phase are explained below, and the efficacy of the data pre-processing is discussed in Section 8.1.

Algorithm 1: Identifying Foot-Ground Contact Period

Input: The image list $ImgLst$, the variance threshold th_{last} from the last session.
Output: The list of foot-ground contact images $FLst$.
for i in range $len(ImgLst)$ **do**
 Calculate the variance of $ImgLst_i$ and save it in list
 $VarLst$;
for i in range $len(VarLst)$ **do**
 if $i = 0$ or $i = len(VarLst) - 1$ **then**
 Continue;
 if $VarLst_i \geq th_{last}$ and $VarLst_{i-1} \geq th_{last}$ and
 $VarLst_{i+1} \geq th_{last}$ **then**
 Center-crop $ImgLst_i$ and save it in $FLst$;
Calculate the top 8% variance value in list $VarLst$ and use it
 to update th_{last} ;
Result: $FLst$

5.1.1 Identifying foot-ground contact periods with variance-based threshold. We observe that the distribution of bright and dark regions in speckle images contains the majority of information about ground surfaces, and that color is not a significant factor. Therefore, to increase the efficiency of our pre-processing phase, we convert all speckle images to grayscale.

The first step, after acquiring the grayscale frames, is to identify speckle images that correspond to the foot-ground contact period. These images are often less blurry, revealing much information about ground surfaces. We note that, when *LaserShoes* is moving in relation to the ground, the collected speckle images are less visible, resulting in lower variances of pixel intensities across an image for the edge of the speckle patterns being fuzzy. Fig. 6 shows some example speckle images from the foot-ground contact period and

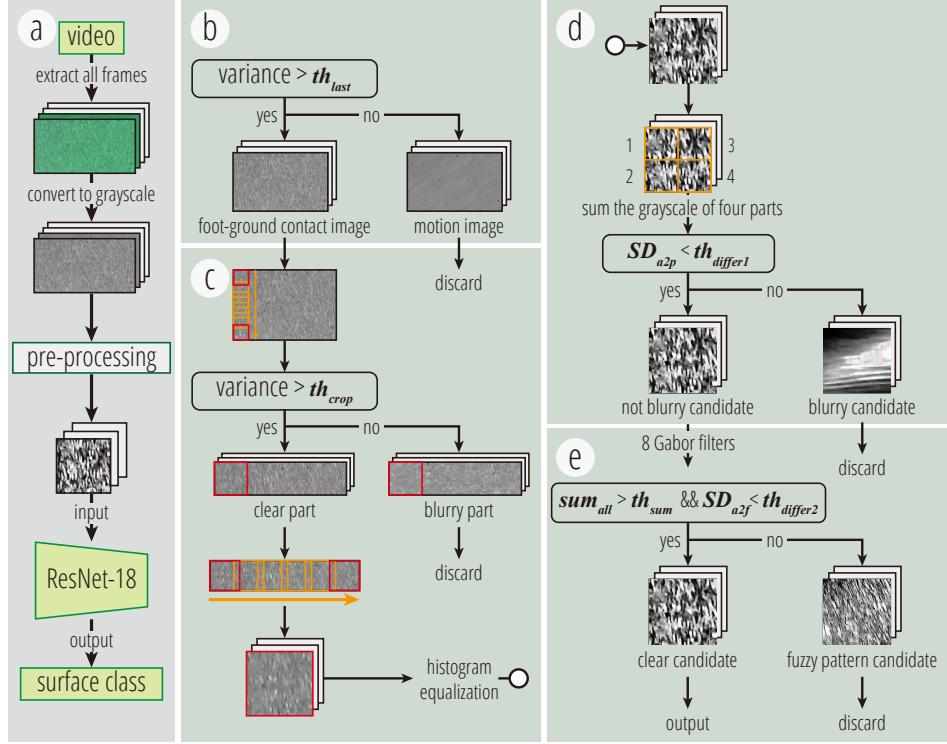


Figure 5: The pipeline of ground surface detection. (a) The overview of this pipeline: all frames of the collected video are converted into grayscale and fed into a pre-processing phase, which identifies a set of images with distinct speckle patterns, which will then be fed into a ResNet-18 model to determine the type of ground surface; (b)-(e) Four main stages of the pre-processing phase; (b) We select speckle images with pixel variances higher than a threshold. These images (i.e., foot-ground contact images) are often captured when the user's foot has solid contact with ground surfaces. Images with low variances are discarded; (c) We crop the foot-ground contact images along the leftmost column and threshold the variance of cropped images. If the cropped image has a higher-than-threshold variance, the entire row of the foot-ground contact images which has the cropped image is preserved for the next steps. Rows with cropped images of lower-than-threshold variances are discarded. Then we slice the preserved rows of the foot-ground contact images to get multiple candidate images which are enhanced for better contrast by histogram equalization; (d) We divide one candidate image into four regions and calculate the sum of each part. If the difference between any two sums is smaller than the threshold $th_{differ1}$, the candidate image will be considered not blurry and passed to the final selection stage; (e) Finally, we apply 8 Gabor filters with various angles to candidate images and calculate the sum of each result. If all the sums are larger than the threshold th_{sum} and differences of any two sums are all smaller than another threshold $th_{differ2}$, the candidate image is clear and ready for the subsequent detection. In other words, these clear candidate images are outputs of this pre-processing phase.

from a user's foot in motion, illustrating the difference in blurriness. Hence, by comparing the variances of pixels, we identify speckle images that are collected from the foot-ground contact period and pass them to the next stage.

We calculate the grayscale variance of each speckle image in each video session. Then, for each speckle image, we recognize it as one collected from the foot-ground contact period if it has a cross-pixel variance that is larger than the top 8% variance value of the previous 90-frame video segment. To further improve robustness, we use adjacent images to aid in identification – we consider a speckle image to be a foot-ground contact image only when both its previous frame and next frame have high variance. Finally, before feeding these selected images into the next stage, we conduct a center crop on them for the lack of sensitivity at the edges of the

CCD module, not being able to output clear laser speckles. The pseudo-code of this pre-processing stage is shown in Algorithm 1.

5.1.2 Cropping images. The first stage yields foot-ground contact images of 1024×592 pixels. We conduct a test to investigate the effect of image size on the detection performance in Section 5.3, and choose 256×256 pixels as the size of our input data. Specifically, we use an extraction window of that size to crop out input images from each foot-ground contact image. This cropping operation also increases the number of samples and improves the efficiency of deep learning model training.

However, within each foot-ground contact image, some regions may still be blurry while others have clear speckle patterns. We eliminate those with blurry speckle patterns in this stage to further

Algorithm 2: Cropping Images

Input: The list of foot-ground contact images $FLst$, the number of cropped images along the height dimension num_h and the number of cropped images along the width dimension num_w .

Output: The list of candidate images $CILst$.

```

for  $i$  in range  $len(FLst)$  do
  for  $j$  in range  $num_h$  do
    Crop the images along height dimension at height index  $j$ ;
    Save the variances of the cropped images in list  $CVarLst$ ;
  Calculate the top 20% of the values in variance list  $CVarLst$  as the threshold  $th_{crop}$ ;
  for  $i$  in range  $len(CVarLst)$  do
    if  $CVarLst_i > th_{crop}$  then
      Calculate the corresponding image index  $idx_i$  and cropped image height index  $idx_h$ ;
      for  $j$  in range  $num_w$  do
        Use the height index  $idx_h$  and the width index  $j$  to crop the image  $F_{idx_i}$  and get  $CandImg$ ;
         $CandImg = \text{HistogramEqualization}(CandImg)$ ;
        Save  $CandImg$  in the list  $CILst$ ;
  Result:  $CILst$ 

```

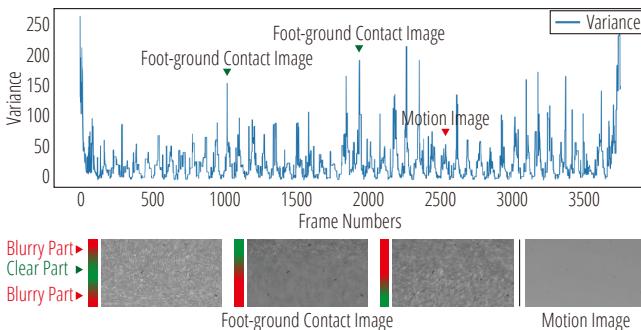


Figure 6: The grayscale variances of all images in one collected video. Images with small variances are motion images that own no speckle patterns but blur while images with large variances correspond to foot-ground contact periods and own speckle patterns. In some foot-ground contact images, some parts own blurry speckle patterns while others own clear ones. And the distribution of speckle patterns is often similar along the image width direction which aligns with the foot breadth direction.

improve our system's robustness. Instead of using the intuitive approach to calculating pixel variance of all cropped images, which could be computationally expensive, we calculate the variances of the cropped images along the left edge of a foot-ground contact image to decide the blurriness of rows in these cropped images reside. We note that the distributions of speckle patterns in each image row are often similar to the rolling shutter of our image

sensor (Fig. 6). Thus, we can determine if a row has clear speckle patterns by inspecting only one section of it. Specifically, we slide the extraction window in the y direction to crop out different image patches and check whether they are clear by thresholding their pixel variances (Fig. 5). The slide stride is 56 pixels, and thus for each foot-ground contact image, six cropped images will be extracted. If a cropped image has a variance higher than the top 20 percent of all variances of all foot-ground contact images belonging to the current video session, we consider it to have clear speckle patterns and save it in a buffer. We also save the indexes of these cropped images for sliding the extraction window along rows of these indexes with a 128-pixel stride. The extracted patches from this step are candidate images. Histogram equalization is applied to candidate images to amplify their contrast. All candidate images are fed into the next pre-processing stage after histogram equalization. Algorithm 2 shows the pseudo-code of this stage.

5.1.3 Removing partial blurry images with region-based sum comparison. There could still be blurry images resulting from the aforementioned stages. To eliminate these images, we design an additional pre-processing stage for fine selection. Because the contrasts of these potentially blurry candidate images become much larger after histogram equalization, the pixel variances of different regions of these images all vary greatly (shown in Fig. 7 (a)). Thus, to identify blurry images, each candidate image is equally divided into four sub-images. We calculate the sum of the grayscale values of every sub-image and eliminate the candidate image if the difference between any two sums exceeds a given threshold. The rest of the candidate images are then fed into the final pre-processing stage.

5.1.4 Removing fuzzy patterns with Gabor filter. Since there may still be relative motions between our sensor and ground surfaces during the foot-ground contact period due to the deformation of ground surfaces, fuzzy patterns can be generated in the speckle images. These fuzzy patterns often appear as stripes oriented in a particular direction, while clear speckle images have patterns with no obvious orientation (as shown in Fig. 7(b) and (c)). To remove images with fuzzy patterns, we apply 8 Gabor Filters with different directions (30, 60, 120, 150, 210, 240, 300, and 330 degrees) and remove those with unbalanced filtered results. Specifically, we eliminate an image if there is a difference between any two filtered results greater than a given threshold. The candidate images that are not eliminated by the third and fourth stages are the output of our pre-processing phase and are the input to the deep learning model. The pseudo-code for these two pre-processing stages is described in Algorithm 3.

5.2 Deep Learning Model

Image classification is a mature field in Computer Vision (CV), and many deep learning algorithms have shown remarkable performance. To choose a proper model for our sensing, we conduct a comparison study with different models, including ResNet-18[21], VGG [47], GoogleNet [55], and MobileNetV3 [24]. As shown in Table 2, ResNet-18 and GoogleNet achieve comparatively high accuracies. We eventually choose ResNet-18 to implement *LaserShoes* for its smaller size, despite its slightly lower accuracy than GoogleNet.

In the ResNet model, input images first pass through a convolution layer, a batch normalization (BN) layer, and a rectified linear

Table 2: Classification accuracy results of different models and their model sizes.

| Model | ResNet-18 | VGG-16 | GoogleNet | MobileNetV3 |
|------------|-----------|--------|-----------|-------------|
| Accuracy | 88.95% | 79.50% | 89.95% | 77.88% |
| Model Size | 42.8M | 512.6M | 48.2M | 6.3M |

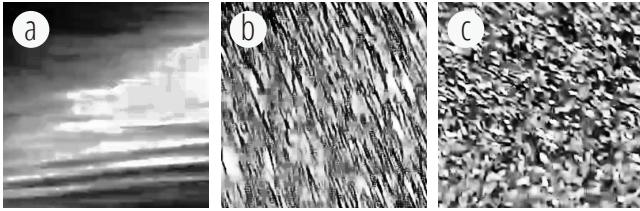


Figure 7: Three kinds of candidate images. (a) Blurry candidate image; (b) Fuzzy candidate image with one-direction stripes; (c) Clear candidate image.

Algorithm 3: Removing Partial Blurry Images and Fuzzy Patterns

Input: The list of candidate images $CILst$, the threshold $th_{differ1}$ which is used to identify partial blurry images, the threshold th_{sum} and $th_{differ2}$ which are used to identify fuzzy patterns and the angle list $ALst$ used for Gabor Filter.

Output: The list of clear candidate images $ClearImgLst$.

for $CandImg$ in $CILst$ **do**

- Divide $CandImg$ into 4 parts and calculate the sum of the grayscale values of every part;
- Calculate the difference between any two sums and save the result in the list $Diff1Lst$;
- if** Every element in $Diff1Lst$ $< th_{differ1}$ **then**
 - for** $Angle$ in $ALst$ **do**
 - $FilterRes = GaborFilter(CandImg, Angle);$
 - Calculate the grayscale sum of $FilterRes$ and save it in the list $SLst$;
 - if** Every sum in list $SLst$ $> th_{sum}$ **then**
 - Calculate the difference between any two sum in list $SLst$ and save the result in the list $Diff2Lst$;
 - if** Every element in $Diff2Lst$ $< th_{differ2}$ **then**
 - Save the $CandImg$ in list $ClearImgLst$;

Result: $ClearImgLst$

unit (ReLU) layer. The data then goes through a series of basic blocks which consists of a residual mapping and an identity mapping. For the residual mapping, the input passes through a convolution layer, a BN layer, a ReLU layer, another convolution layer, and another BN layer, while for the identity mapping, the input only passes through a 1×1 convolution layer to be downsampled to the same size as the residual mapping result. Then the two mapping results are added and the sum passes through a ReLU layer to get the output of a

basic block. Finally, an average pooling and a full connection layer are operated to obtain the classification results. During training, we select Cross Entropy Loss as the loss function and use the Adam optimizer. The learning rate and the batch size are set to 0.0001 and 32, respectively. We do not use a pre-trained model to initialize our parameters and use 150 epochs for the model training because we find that it is enough for our models to be converged.

5.3 Image Size Selection

The model's input is the clear candidate images from the data pre-processing phase, and the model's output is the type of ground surfaces. Image size is set to 256×256 in our ground surface detection, the same as the size used in SensiCut [13]. To verify the efficacy of this image size, we extract a number of clear candidate images with different sizes to train a series of ResNet-18 models. The experimented image sizes included 64×64 , 128×128 , 256×256 , and 512×512 . The results of average accuracy and inference time for the classification of one input image are shown in Table 3. As expected, input images with larger sizes lead to higher accuracy but take significantly longer to classify. Given the improvement in accuracy is modest from 256×256 to 512×512 , we select 256×256 as the size of the input images to our model to balance accuracy with inference time.

Table 3: Classification accuracy and inference time for one image with various input image sizes.

| Image Size | 64×64 | 128×128 | 256×256 | 512×512 |
|----------------|----------------|------------------|------------------|------------------|
| Accuracy | 48.33% | 66.95% | 88.95% | 94.67% |
| Inference Time | 2ms | 4ms | 17ms | 45ms |

5.4 Real-Time Inference

In real-time detection, the image sensor continually records frames, and every 90 frames constitute a video session that is fed into the pre-processing stage. If no clear candidate images are detected by the pre-processing phase, the detection pipeline outputs “None” as a neutral label. We conduct testing using 100 video sessions captured during participants' normal walks on various everyday ground surfaces. Our result shows that for every video session, after the data pre-process phase, the average number of input images fed into the subsequent model is 11. We use C++ for implementing the data pre-processing for a superior speed and use Python for implementing the deep learning model. For every input image of a video session, the classification model will output a corresponding surface type. Among all these types, we choose the surface type that appears the most frequently as the surface label of this video session. And the label is also provided to the user as the detection feedback. We record the average time needed for completing the pre-processing and inference of one video session, with 100 sessions collected from various participants and ground surfaces processed on a Raspberry Pi Zero 2 W, a laptop with a CPU of 3.1 GHz dual-core Intel Core i5, and a GPU of NVIDIA GeForce RTX 3090 respectively. Results are shown in Table 4. We find that the current implementation of *LaserShoes* running solely on the Raspberry Pi board cannot

perform real-time detection without dropping input images if the duty cycle of users' feet contacting ground surfaces is too high, which we acknowledge as a limitation of our system.

Table 4: Data processing pipeline average run time for one session on various devices.

| Model | Pre-processing | Inference | Total |
|-----------------|----------------|-----------|--------|
| Laptop CPU | 99ms | 1211ms | 1310ms |
| Embedded System | 696ms | 6082ms | 6778ms |
| GPU | 75ms | 194ms | 269ms |

6 EVALUATION

Our user study consisted of one main study and three supplementary investigations. The main study involved collecting data with 24 ground surfaces to understand *LaserShoes*' ability to classify the ground surface material while its wearer is walking. In the supplementary studies, we aimed to evaluate the robustness of *LaserShoes* under various conditions (i.e., on dry, wet, and icy surfaces, on sand surfaces of different grain sizes, and under different lighting conditions).

Considering that when pre-processing, identifying the foot-ground contact periods of a 1.5s video session is the first stage and is the basis of the subsequent pre-processing stages, a high detection accuracy (DA) of identifying the foot-ground contact period (FGCP) is necessary. Thus, we first evaluated this detection accuracy, which is defined as

$$DA = \frac{\# \text{ detected 1.5s video sessions containing FGCP}}{\# \text{ all 1.5s video sessions containing FGCP}}.$$

Then, we used accuracy, precision, recall, and F1 score as our evaluation metrics for the ground surface classification. To calculate them, we only considered the 1.5s video sessions that have surface label (SL) output and eliminated those with "None" signals. The classification accuracy (CA) is defined as

$$CA = \frac{\# \text{ correctly classified 1.5s video sessions with SL output}}{\# \text{ all 1.5s video sessions with SL output}}.$$

6.1 Main Study with 24 Ground Surface

6.1.1 Ground surface materials. We selected a total of 24 common ground surfaces, comprising 15 indoor surfaces and 9 outdoor surfaces, for our study. These surfaces could be classified into five groups: 1) rough, 2) smooth, 3) hard, 4) discontinuous, and 5) granular. These surfaces are shown in detail in Fig. 8. For each ground surface, we prepared at least one continuous area of 20 square meters in size to allow our participants to walk naturally (e.g., not need to frequently turn or turn back, not need to keep looking down the ground) during data collection for our study.

6.1.2 Participants and apparatus. We recruited 15 participants (7 males and 8 females), with ages ranging from 20 to 27 years old (mean = 23.40, SD = 1.56) via social media and flyers. Their body weights ranged from 48.0kg to 82.6kg (mean = 61.03, SD = 9.93) and their heights ranged from 158.5cm to 182.0cm (mean = 170.13, SD = 6.83). Of all the participants, 5 wore sneakers, 6 wore running shoes, 3 wore canvas shoes, 1 wore ankle boots, and 1 wore snow

boots. Their shoe sizes ranged from 23.0cm to 27.0cm, with a mean of 24.67 (SD = 1.12).

Participants wore their own shoes normally and our *LaserShoes* as described in Section 4 to collect videos from ground surfaces while participants were walking on them. Considering that our device requires proximity to ground surfaces, we required participants to wear flat shoes. Fig. 9 shows some example shoe styles that *LaserShoes* is compatible with. Distances between our image sensor and ground surfaces in the study varied from 6cm to 10cm across the 15 participants. The detection component was attached tightly to participants' shoes through our designed clamping mechanism, while the processing and assistant component was attached to participants' lower legs using Nylon tapes.

6.1.3 Data collection procedure. We started the study with an introduction of the procedure and helped the participant put the devices on. For each surface, we used tapes to indicate an area that the participants could walk on. Participants were allowed to walk freely in the area. Each study had two sessions. A short practice session was at the start, where the participant walked through all surfaces. This session was used to familiarize participants with the system and no data was collected. We asked the participant to slow down their walk if no clear speckle patterns could be captured by *LaserShoes* (i.e., output from the pre-processing phase). After the practice session, the participants were asked to walk on each chosen surface for 1~2 minutes in the second session for data collection. The order of the surfaces each participant needed to walk on was randomized to avoid bias (e.g., a change in walking speed or gait caused by fatigue). In addition, in order to simulate real-world scenarios, participants were asked to adjust their *LaserShoes* after each session and to take breaks in between sessions (around 2 mins). The study was conducted under typical indoor and outdoor lighting conditions. To collect the ground truth of foot-ground contact periods, a camera was set up to record the foot movements of participants during the study and research assistants labeled all foot-ground contact timestamps manually. In total, we collected 28,492 1.5s video sessions on 24 surfaces from the 15 participants. And it took around 2 hours for each participant to finish the data collection.

6.1.4 Results. To evaluate the performance of our system for ground surface classification, we used both within-user and cross-user approaches. For within-user evaluation, to ensure there is no overlapping between the training set and the test set, we first split all data into ten folders and randomly selected two folders as the test datasets. Of note that no time-adjacent input images were included in both training or test datasets. For cross-user evaluation, we used leave-one-out evaluation methods using 14 participants' data to train and the remaining one to test.

Detection Accuracy of Identifying Foot-Ground Contact Periods. The collected videos were processed using the method described in Section 5.1 and we first evaluated the performance of identifying foot-ground contact periods using the formula defined above. The detection accuracy is 90.91%, indicating that our method can detect the majority of foot-ground contact periods from recorded data.

Within-User Evaluation Results. Results of the within-user detection accuracy for 24 ground surfaces are shown in Fig. 10 (a).



Figure 8: 24 kinds of ground surfaces that were selected for the user study. 15 of them are indoor, while the other 9 surfaces are outdoor surfaces. Based on their characteristics, these ground surfaces are divided into five categories: rough, smooth, hard, discontinuous, and granular.



Figure 9: Five examples of how *LaserShoes* can be worn on shoes of different styles: (a) Snow boots; (b) Ankle boots; (c) Running shoes; (d) Canvas shoes; (e) Sneakers.

The average classification accuracy of the 24 ground surfaces is 86.93%, with the recall of 87.17% ($SD = 10.09$), the precision of 85.82% ($SD = 13.57$) and the F1 score of 85.94% ($SD = 10.59$). For 15 indoor surfaces, the average classification accuracy is 91.53%, with the recall of 90.60% ($SD = 9.62$), the precision of 92.48% ($SD = 7.23$) and the F1 score of 91.23% ($SD = 7.06$), while for 9 outdoor surfaces, the average classification accuracy is 78.86%, with the recall of 81.46% ($SD = 8.07$), the precision of 74.73% ($SD = 14.39$) and the F1 score of 77.13% ($SD = 9.58$). Indoor surface detection is more accurate than outdoor surface detection. The reason for this could be that the light condition outside is less stable than it is indoors due to changes in intensity and angle of sunlight. This may reduce the quality of collected images, resulting in poor detection results.

Besides, we also evaluated the detection accuracy of surfaces with different characteristics and the results are shown in Table 5. The results show that rough surfaces have the highest accuracy and the lowest standard deviation among the five surface groups with varying characteristics. This makes sense because the microstructure of rough surfaces is more complex, resulting in more subtle patterns. Furthermore, discontinuous surfaces have the lowest average accuracy and a large standard deviation.

Cross-User Evaluation Results. For cross-user evaluation, the detection results are shown in Fig. 10 (b). The average classification accuracy of the cross-user model is 80.57%, with the recall

of 80.36% ($SD = 10.48$), the precision of 78.32% ($SD = 17.62$) and the F1 score of 78.73% ($SD = 13.86$). For indoors and outdoors, the average classification accuracy are 83.22% and 73.13%, with the recalls of 85.48% ($SD = 8.95$) and 71.85% ($SD = 6.56$), the precision of 86.79% ($SD = 10.00$) and 62.54% ($SD = 16.21$), and the F1 scores of 86.39% ($SD = 8.45$) and 65.97% ($SD = 11.53$), respectively. In contrast to within-user results, classification accuracy decreases in the cross-user evaluation. This could be due to the fact that participants were wearing different shoes in the study, which caused different distances between the image sensor and ground surfaces. Furthermore, different foot postures of participants when their feet come into contact with ground surfaces contribute to a decrease in accuracy. Some participants' feet were in aversion, while others were in inversion or in neutral positions. These different foot postures (shown in Fig. 11) cause a distance change between the image sensor and ground surfaces. The distance differences result in differently formed speckle patterns and thus variance between training and test datasets – the same type of ground surface may correspond to multiple speckle patterns. This variance may decrease the accuracy of the cross-user evaluation. And indoor detection, like within-user results, outperformed outdoor detection.

We also tested the performance of the cross-user model for five groups of surfaces with different characteristics. The results are

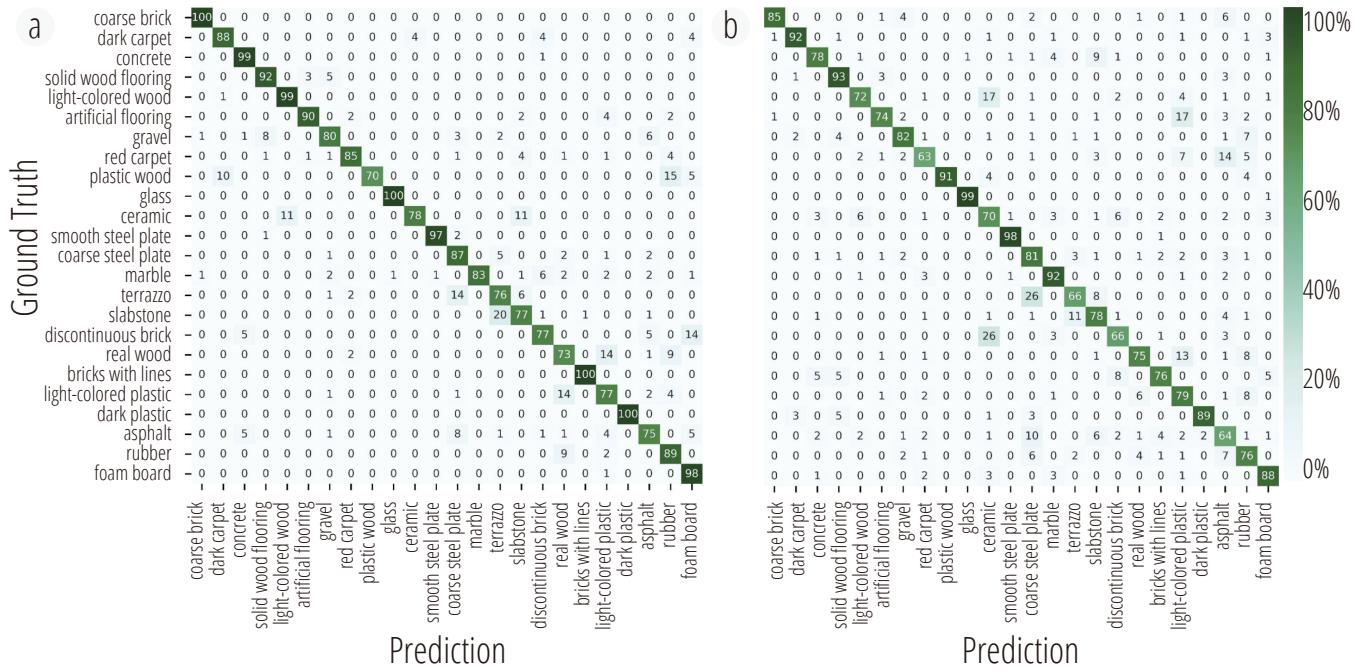


Figure 10: The confusion matrices of two trained classification models of the 24 ground surfaces. (a) Classification results using a within-user model. (b) Classification results using a cross-user model.

Table 5: Average Accuracy (%) and SD results of the within-user model and cross-user model for five surface characteristics.

| | Rough | Smooth | Hard | Discontinuous | Granular |
|-------------|--------------|---------------|---------------|---------------|--------------|
| Within-user | 91.77 ± 6.92 | 86.57 ± 12.98 | 84.03 ± 7.66 | 83.40 ± 11.87 | 85.33 ± 9.91 |
| Cross-user | 79.93 ± 9.67 | 87.04 ± 10.79 | 82.98 ± 11.00 | 72.42 ± 4.41 | 77.24 ± 8.87 |

Table 6: Ground surface classification results in different lighting conditions.

| Lighting Conditions | Indoor-with-light | Indoor-without-light | Outdoor-at-daytime | Outdoor-at-dusk | Outdoor-at-night |
|---------------------|-------------------|----------------------|--------------------|-----------------|------------------|
| Accuracy (%) | 90.05 | 88.99 | 71.85 | 90.94 | 87.69 |

shown in Table 5, which indicates that compared to within-user results, detection accuracy did not change a lot for smooth and hard surfaces. However, for rough, discontinuous, and granular surfaces, there is a large decrease. The reason may be that surfaces with

complex microstructure amplified the difference in participants' foot postures, resulting in larger differences of speckle patterns belonging to the same type of ground surfaces.

Visually Similar Ground Surfaces. Among our selected ground surfaces, light-colored wood and artificial flooring look very similar, which are not easy to distinguish by conventional RGB cameras. The results shown in Fig. 10 reveal that in both within-user and cross-user conditions, these two visually similar surfaces can be distinguished from the other one with *LaserShoes*.

6.2 Supplementary Investigation

Given the length of the primary data collection, the supplementary study is not conducted on the same day to avoid the fatigue of participants. 12 participants took part in our supplementary studies. The basic procedure was the same as the main study procedure. We

Figure 11: Three types of foot postures. (a) Eversion; (b) Neutrality; (c) Inversion.



finally collected 19,319, 4,250, and, 41,005 1.5s video sessions for each study, respectively.

6.2.1 Dry, wet, and icy surfaces. In outdoor settings, ground surfaces could be dry, wet, or icy due to different types of weather. This may pose a potential danger to pedestrians. Thus, the sensing capability of *LaserShoes* to identify ground surface conditions could have real-world uses. We conducted experiments to classify ground surface conditions on nine types of outdoor surfaces, shown in Fig. 8, under three conditions (i.e, dry, wet, and icy). For the wet condition, we poured water on the ground while for the icy condition, we put crushed ice on the ground. We conducted two evaluations in this study. We treated each combination of surface and condition as a separate label (27 in total) in the first validation. In the second evaluation, we combined all of the surfaces of the icy condition into one label (19 in total). The detailed results are shown in Fig. 12. In the first evaluation, the detection model has a 62.89% recall, a 66.06% precision, and a 59.91% F1. In the second evaluation, after merging icy surfaces, the detection model has a 76.06% recall, a 76.75% precision, and a 74.29% F1. These results show the feasibility of *LaserShoes* detecting ground surface conditions in real-world applications to improve pedestrian safety.

6.2.2 Sand surfaces with different grain sizes. Even when the material is the same, the physical state of the material (e.g., graininess, looseness) can vary. We also investigated how *LaserShoes* could perform finer-grained ground surface material sensing. Participants were asked to walk on three different types of sand surfaces with the same procedure as the main study. To be more specific, we assess the classification performance using data collected on sand surfaces with sands of three different grain sizes (i.e., small, medium, and large). The classification accuracy for the sand types is 92.28% with an 87.60% recall, a 95.56% precision, and a 90.59% F1, which indicates that *LaserShoes* could identify the same type of surfaces with different fine-grained surface geometries.

6.2.3 Different lighting conditions. Lighting conditions may affect the quality of speckle images and thus the ground surface detection performance. To test the robustness of *LaserShoes* against this factor, we collected data in five different lighting conditions. These conditions included two for the 15 indoor surfaces and three for the 9 outdoor surfaces, and are listed as follows:

- Indoor-with-light: lamps (cold light source) on in a room.
- Indoor-without-light: lamps off in a room.
- Outdoor-at-daytime: much sunlight outdoors at daytime.
- Outdoor-at-dusk: little sunlight outdoors at dusk.
- Outdoor-at-night: no sunlight, with streetlamps on, outdoors at night.

We trained five classification models, each using the data collected under different lighting conditions. Table 6 shows the average surface classification accuracies for the five different lighting conditions. The results demonstrate that, with the exception of the outdoor-at-daytime condition, the classification accuracy for all other conditions was above 87%. This indicates the robustness of *LaserShoes*, except under lighting conditions with strong ambient light, which requires further improvement.

7 APPLICATION

To demonstrate our system as a real-time assistant in many use cases by sensing ground surfaces, we developed five application examples as shown in Fig. 14.

7.1 Personal Running Assistant

A considerable amount of research has been dedicated to assisting and promoting running activity. For instance, sensing techniques have been developed to help users understand their body (e.g., tracking kinesiological data about feet and gait) [59], some data-driven interfaces are designed to motivate users' actions [34], while others are proposed to support natural navigation running in unknown places [28, 46]. Some previous works have taken the form of smart shoes, which people envision as being capable of adapting to different terrains to improve runner performance and health in the future, becoming an active support tool [33].

However, there are currently few smart shoes that can yield rich terrain surface information that one can use to correlate with running experience. For example, a cross-country runner who runs over a variety of ground surfaces of varying difficulty levels may want to understand how running performance is related to the ground surface. Our body has different reactions and biomechanical demands with different types of ground [3, 15]. For instance, the degree of compliance of the ground surface will impact the speed of energy transfer between people's foot and ground surface, resulting in different foot-ground contact time and energy consumption [23, 32]. In this case, *LaserShoes* could be used to support running analysis and yield guidance with fine granularity. Fig. 14 (a) shows an example of using *LaserShoes* to support running analysis. During the running trial, the user ran over various ground surfaces such as carpet, rubber, asphalt, and discontinuous brick, and *LaserShoes* detected these different surfaces. As a result, the detection results could be used to generate reports for each surface, such as time, speed, and energy consumption.

7.2 Gait Analysis

Gait parameters variability is an important diagnostic indicator of health [41], related to both the quality of life and mortality [51], correlating with the rehabilitation degree of specific joint injuries [56], and thus has received significant attention to both clinicians and researchers. However, the terrain type can significantly influence the gait pattern [31, 50], which underscores the need to consider different terrain types while analyzing. Our *LaserShoes* can be used to support such analysis. Specifically, as shown in Fig. 14 (b), when the user steps on soft surfaces like sand and mud, her gait will change due to the softness of the surface. However, when stepping on hard surfaces like asphalt, the user can maintain a normal gait. We can incorporate a simple IMU module into our *LaserShoes* to monitor users' gait information, as well as use *LaserShoes* to collect terrain ground surface information. In this case, the additional information can be leveraged to examine how gait is changed on various types of ground surfaces, providing insights that could be of use in medical applications.

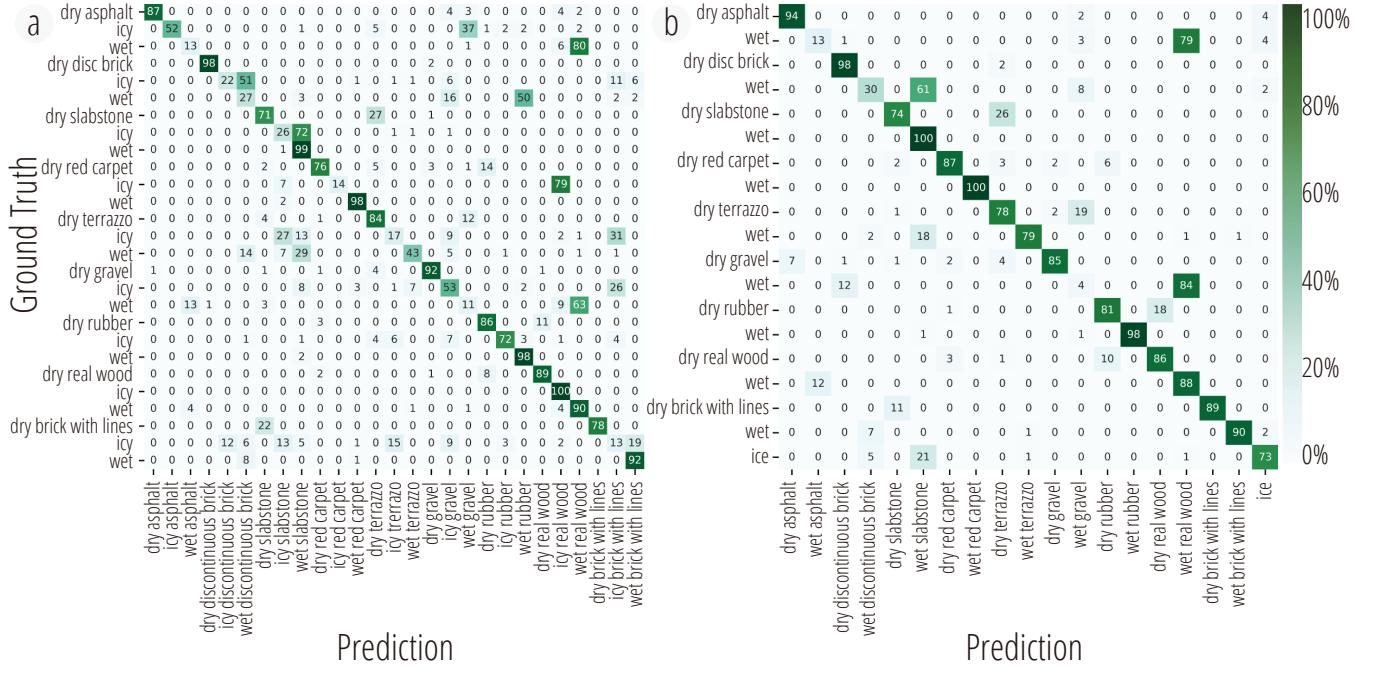


Figure 12: The confusion matrices of two trained classification models to identify dry, wet, and icy ground surfaces. (a) Classification results using the model that differentiates various icy ground surfaces; (b) Classification results using the model that merges various icy ground surfaces into one category.



Figure 13: Five kinds of lighting conditions in which we collected data. (a) Indoor-with-light; (b) Indoor-without-light; (c) Outdoor-at-daytime; (d) Outdoor-at-dusk; (e) Outdoor-at-night.

7.3 Cleaning Equipment Auto-Control

There is a wide variety of cleaning equipment (e.g., UnoClean¹), designed for indoor and outdoor applications. Many advanced cleaning machines have numerous cleaning modes (e.g., vacuuming power, whether water is used) for various types of ground surface and dirt cleaning needs. For example, the cleaning mode used for grass cannot be directly applied to real leather carpets. A high-power mode will likely damage the carpet. In this case, users must frequently change the working mode due to the various physical forms and chemical compositions of the ground surface. As the variety of decoration materials in our living environments grows, automatically switching the cleaning machine's working mode based on the floor material can provide much convenience and reduce errors in our daily cleaning tasks. As shown in Fig. 14 (c), if

the user wears *LaserShoes* while cleaning, our system detects the material of the floor the user is walking on, such as ceramic, carpet, or wood, and automatically changes the cleaning mode of the machine. Similar to floor cleaning equipment, other types of mobile tools (e.g., pressure washer, leaf blower) or even smart devices (e.g., smartphones, AR/VR headsets) could also leverage ground surface as side-channel information to improve their performances.

7.4 Coarse Navigation

Navigation tools have greatly facilitated our lives. Even with GPS navigation, people might get disoriented in outdoor places with complex layouts or crowded areas. GPS also does not work in indoor settings such as museums, airports, and shopping malls [5]. These environments often have floors made of various materials. For example, different stores in the mall may have different decorative floor materials. The route for outdoor running may include grass,

¹<https://www.unoclean.com/>

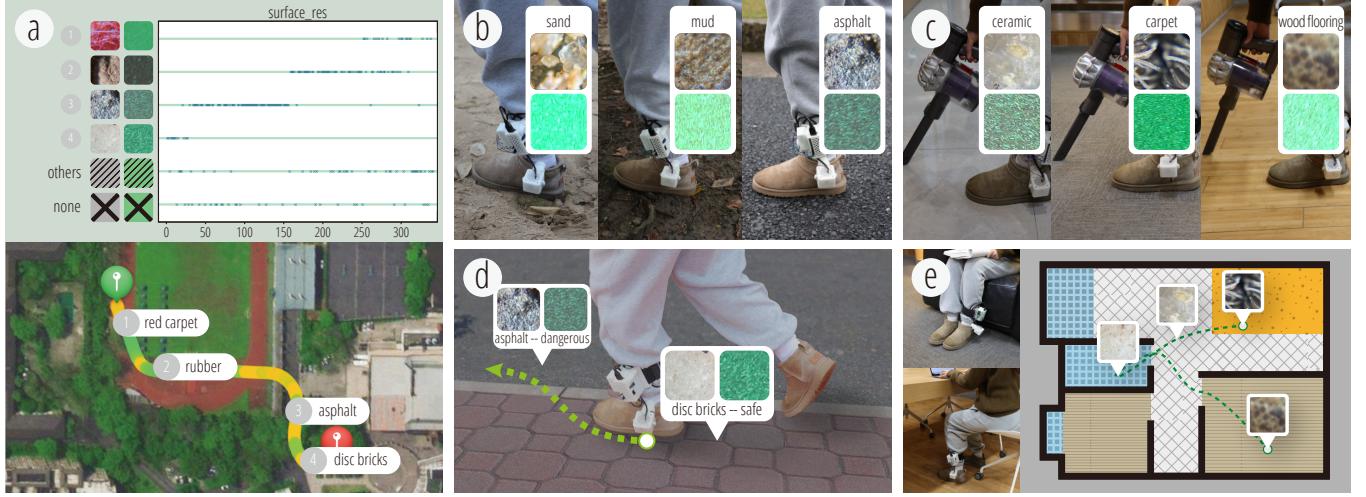


Figure 14: Five applications using *LaserShoes*. (a) Personal running assistant: *LaserShoes* can detect ground surface that users are currently running on, and these detection results can be used to generate running analysis reports for each surface; (b) Gait analysis on different terrains: When combined with gait analysis sensors (e.g., IMU), *LaserShoes* can help users to detect changes in gait on different terrains; (c) Cleaning equipment auto-control: When a user is cleaning, *LaserShoes* will detect on which surface the user is stepping and the working mode of the cleaning equipment can be changed automatically according to the detection feedback; (d) Coarse navigation: *LaserShoes* can provide users coarse navigation. For example, when the detection feedback of *LaserShoes* is brick, it means that the user is walking in the proper way. However, if the feedback changes to asphalt that is unexpected, it means the user may be walking in the wrong, or a dangerous way, and some alerts will be given to the user; (e) Daily activity recognition through localization: The space in which a user is staying can be recognized by using *LaserShoes* to detect ground surface types. For instance, detecting carpet is likely to correspond to staying in the living room for entertainment, while detecting wood may mean staying in the study and working. Thus, based on space recognition, we can calculate how much time the user spends in various spaces and roughly achieve an analysis of the user's daily activities.

gravel, asphalt, and other surfaces. *LaserShoes* can infer coarse user locations from ground surfaces and alert users when they are off-course. As shown in Fig. 14 (d), the proper route for the user is the sidewalk made of bricks. However, if *LaserShoes* detects that the current surface is asphalt, its user is likely on the wrong route and will receive an alert.

The negation system can also be applied to accessibility for which we envision *LaserShoes* to work in concert with accessible infrastructure in urban environments. Visually impaired individuals could rely on additional information (e.g., tactile feedback of ground surfaces) to acquire spatial awareness [63]. Previous research attempted to design physical tactile maps to enable users to access information with audio [22, 27, 44, 52], tactile [58], and a combination of tactile and audio feedback [18, 25, 39, 64]. Instead of relying solely on the tactile sensation of users' feet (e.g., tactile ground surfaces, blind pathways), *LaserShoes* could sense ground surfaces for users, providing an alternative solution that could take advantage of sensory substitution techniques – converting ground textures into sounds to guide visually impaired individuals to stay on track of pathways that are safe for them.

7.5 Daily Activity Recognition through Localization

Recognized activities provide rich contextual information to support natural human-computer interactions. Statistical analysis of a

person's behavior in an environmental space helps with the inference for the design of the space and the user's lifestyle. For instance, logged activity data can be used to help older adults to encourage healthy daily routines and active lifestyles, and to monitor chronic health conditions and enjoyment [29]. Among all types of in-home contextual information, the ground surface texture is often unique to living spaces of different functions. For example, the bathroom floor is typically made of easy-to-clean and waterproof tile surfaces, the bedroom floor soft carpets or rugs, and the living room wood or plastic floor materials. In this case, we can use *LaserShoes* to recognize the ground texture and determine which space the user is in to coarsely infer their activities. As shown in Fig. 14 (e), when *LaserShoes* detects carpet, the user is more likely to be relaxing in the living room. However, when *LaserShoes* detects wood floors, the user is more likely to be working in the study. We can, for example, alert users when they spend too much time on the toilet (e.g., playing with smartphones), which is detrimental to their health.

8 DISCUSSION

8.1 Efficacy of Data Pre-processing

Although machine learning models are somewhat resilient to noisy data points, they require more computation power during inference. To alleviate the such burden, a denoising process is commonly performed prior to feeding into machine learning models [68, 69]. In our case, if we do not remove blurry images, the time consumption

for inference will be large, which is opposite to our goal of real-time prediction. Even if we only extract one image by cropping one raw frame and do not perform data pre-processing, the number of images from one video session fed into the classification model will be 90. However, the average number of images fed into the classification model after data pre-processing is 11, indicating that our data pre-processing step can significantly reduce computation costs during inference.

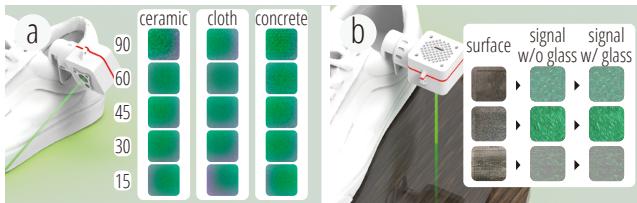


Figure 15: (a) Illustration of early alert by sensing ground surfaces ahead of users. When facing toward the front with the angles of 60, 45, 30, and 15 degree, *LaserShoes* could still capture discernable speckle patterns; (b) Example speckle images captured on three different surfaces with and without a transparent glass coating layer.

Further, to evaluate the influence of the data pre-processing step in terms of ground surface classification performance, we conducted experiments on data collected from one of our participants. The experiment procedure is the same as our main study except that we replaced the pre-processing part with cropping one 256×256 image from each frame. For the classification model trained with raw data, the recall, precision, and F1 are 64.25%, 67.22%, and 60.60%, respectively. For the classification model trained with data after pre-processing, the recall, precision, and F1 are 88.45%, 88.05%, and 87.60%, respectively. Therefore, conducting our data pre-processing step can achieve better performance compared to using raw data.

8.2 Avoid Overfitting

Overfitting is a common issue in deep learning applications, especially when the number of training samples is small. To prevent the deep model in our system from overfitting, common techniques, including data augmentation, and normalization, were applied during the training process. Besides, as described in Section 5.1.2, cropping a raw speckle image to generate multiple smaller input images helps increase the number of training samples. Moreover, we set the number of training epochs of the model to 150, after we conducted experiments using a validation set of data and found that the training loss converged while the validation loss did not degrade at around 150 epochs. The evaluation results with high classification accuracies, especially those from the cross-user study, demonstrate effective mitigation of overfitting.

8.3 Power Consumption

There are three main parts that consume power in our system: The laser emitter with a switch module (51.3 mW), the image sensor (1047.9 mW), and the Raspberry Pi (2643.6 mW). *LaserShoes* has a relatively high total consumption, which prevents it from being

continuously used for a long time without battery exchange. In the future, to reduce the power consumption on Raspberry Pi, the collected data could be transferred to a cloud server via low-power wireless communications. We could also design a custom circuit and reduce power consumption by removing components that are not in use and using low-power MCU and communication modules. Besides, the current image sensor captures images of 1280×720 pixels for efficient data collection. However, in live classification, the input images need not be that large, possibly taken by smaller image sensors to preserve power.

8.4 Sense Surfaces ahead for Early Alert

Since *LaserShoes* uses images captured when a user's foot is in contact with the ground, our system in its current implementation could not predict ground surface conditions in advance, limiting use scenarios such as alerts of dangerous surface conditions. To achieve this, *LaserShoes* should be able to leverage in-flight images. To mitigate the motion effect, we could add an IMU sensor to measure motion speed and implement deblurring methods [9, 17]. Image sensors with a short exposure time could also help to obtain clear images when the user's foot is moving in the air. Second, we could tilt up our device to sense ground surfaces in front of a user for early alerts (Fig. 15 (a)). We performed a test to see if our sensing system could still function with our device tilted up, pointing to the front of a shoe. Results indicate discernable speckle patterns up to 45 degrees for the three types of surfaces we tested (Fig. 15 (b)). However, it merits future research to investigate how this sensor configuration could work in real-use cases powered by real-time signal processing and classification.

8.5 Loose or Transparent Ground Surfaces

In practice, we discover that *LaserShoes* cannot capture frames with high-quality speckle patterns on loose ground surfaces such as grass for insufficient reflected light intensity. We suspect that grass surfaces diffused or absorbed most of the laser energy due to their layered surface micro geometries. Besides, for transparent ground surfaces such as glass (Fig. 15 (b)), the reflected laser is also weakened. Though speckle patterns can still be formed on transparent ground surfaces, information on the textured surfaces underneath the transparent coating layer is much deluded, resulting in less discernable speckle patterns than ones induced on surfaces without the transparent coating layer.

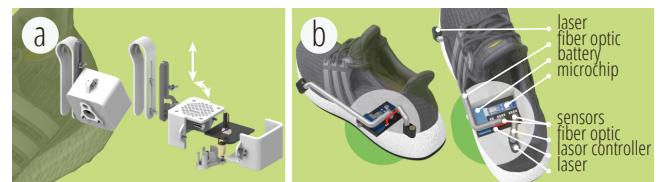


Figure 16: Illustration of two alternative designs with optimized form factors. (a) Attaching the system onto a height-adjustable mechanical module; (b) Integrating the system into a smart sole.

8.6 LaserShoes under Intense Ambient Light

When the ambient light is too intense, the image sensor receives too much ambient light, which lowers the signal-to-noise ratio (SNR). As a result, the speckle patterns become blurry or invisible under some outdoor conditions in our study. To mitigate this issue, future systems could leverage optical filters. Given that laser light is polarized and has a narrow frequency band, we could include a polarizer or a band-pass filter between ground surfaces and the image sensor. These filters could make the laser a dominant signal on captured laser speckle images that have sufficient SNR for classification. Another tactic to preserve SNR is to implement synchronous detection, with the image sensor and the laser in sync. Specifically, we could leverage high-speed image sensors to take two consecutive photos with and without the laser turning on. The subtraction between these two consecutive photos should reveal little effect imposed by the ambient light which is relatively constant and therefore the effect could be subtracted out.

8.7 Form Factor Optimization

Our current implementation is relatively bulky. Furthermore, different image sensor heights, which are affected by shoe styles and foot postures, will reduce ground detection accuracy as discussed in Section 6.1.4. In the future, *LaserShoes* could be replicated with better form factor designs.

One possible solution is to make the height of the image sensor consistent across shoe styles by adding a height-adjustable mechanical module as shown in Fig. 16 (a). This module could also mitigate variances introduced by the foot posture by asking users to calibrate and adjust *LaserShoes* before use.

Since the diode of a laser emitter and the chip of an image sensor are both very small, they can be combined into a single integrated component that might be sufficiently thin to be integrated on a smart sole under shoes as shown in Fig. 16 (b). In this case, the sensing distance is short and consistent, and the sensor is isolated from the ambient light when the sole is in contact with ground surfaces, all of which could result in an improved SNR that yields higher classification accuracies.

9 CONCLUSION

We present *LaserShoes*, a texture-sensing wearable system that detects ground surfaces using Laser Speckle Imaging. Our system can retrofit shoes, and consists of a laser emitter that illuminates ground surfaces and an image sensor that records videos with laser speckles. The recorded videos first pass through a pre-processing phase with which we extract input images from speckle images captured when a user's foot is in contact with ground surfaces. Next, these input images are fed into a ResNet-18 classification model for surface type detection. We conducted a main study and three supplementary investigations to evaluate our system's classification accuracy and robustness across various surface conditions and under different lighting conditions. We showed five applications of *LaserShoes* to demonstrate its potential use cases. Finally, we discussed our evaluation results and future work needed to further improve our system.

ACKNOWLEDGMENTS

This project is supported by the National Natural Science Foundation of China (No. 62202423), and the Fundamental Research Funds for the Central Universities (No. 2022FZZX01-22). We thank all study participants for participating in the study and reviewers for constructive comments.

REFERENCES

- [1] Paramvir Bahl, Venkata N. Padmanabhan, Venkat Padmanabhan, and Victor Bahl. 1999. *User Location and Tracking in an In-Building Radio Network*. Technical Report MSR-TR-99-12. Citeseer. 12 pages. <https://www.microsoft.com/en-us/research/publication/user-location-and-tracking-in-an-in-building-radio-network/>
- [2] Nagaraj N. Bhat, Samik Dutta, Surjya K. Pal, and Srikanta Pal. 2016. Tool condition classification in turning process using hidden Markov model based on texture analysis of machined surface images. *Measurement* 90 (2016), 500–509. <https://doi.org/10.1016/j.measurement.2016.05.022>
- [3] Frans Bosch and K Klomp. 2005. *Biomechanics and exercise physiology applied in practice*. Elsevier Churchill Livingstone, Churchill Livingstone, London, United Kingdom.
- [4] Justin Chan and Shyamnath Gollakota. 2022. Laser Speckle Using Smartphone LiDAR. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services* (Portland, Oregon) (*MobiSys '22*). ACM, New York, NY, USA, 632–633. <https://doi.org/10.1145/3498361.3538670>
- [5] Hsuan Hsuan Chang. 2015. Which one helps tourists most? Perspectives of international tourists using different navigation aids. *Tourism Geographies* 17, 3 (2015), 350–369.
- [6] Jianshe Chen. 2007. Surface Texture of Foods: Perception and Characterization. *Critical Reviews in Food Science and Nutrition* 47, 6 (2007), 583–598. <https://doi.org/10.1080/10408390600919031> PMID: 17653982.
- [7] Xiang 'Anthony' Chen, Julia Schwarz, Chris Harrison, Jennifer Mankoff, and Scott E. Hudson. 2014. Air+touch: Interweaving Touch & in-Air Gestures. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) (*UIST '14*). ACM, New York, NY, USA, 519–525. <https://doi.org/10.1145/2642918.2647392>
- [8] Jingyuan Cheng, Oliver Amft, Gernot Bahle, and Paul Lukowicz. 2013. Designing sensitive wearable capacitive sensors for activity recognition. *IEEE sensors journal* 13, 10 (2013), 3935–3947.
- [9] Sunghyun Cho and Seungyong Lee. 2009. Fast Motion Deblurring. *ACM Trans. Graph.* 28, 5 (dec 2009), 1–8. <https://doi.org/10.1145/1618452.1618491>
- [10] Bernard Choi, Nicole M. Kang, and J. Stuart Nelson. 2004. Laser speckle imaging for monitoring blood flow dynamics in the *in vivo* rodent dorsal skin fold model. *Microvascular Research* 68, 2 (2004), 143–146. <https://doi.org/10.1016/j.mvr.2004.04.003>
- [11] Bernard Choi, Julio C. Ramirez-San-Juan, Justin Lofti, and J. Stuart Nelson M.D. 2006. Linear response range characterization and *in vivo* application of laser speckle imaging of blood flow dynamics. *Journal of Biomedical Optics* 11, 4 (2006), 041129. <https://doi.org/10.1117/1.2341196>
- [12] Sungwoo Chun, Yeonhai Choi, Dong Ik Suh, Gi Yoon Bae, Sangil Hyun, and Wanju Park. 2017. A tactile sensor using single layer graphene for surface texture recognition. *Nanoscale* 9, 29 (2017), 10248–10255.
- [13] Mustafa Doga Dogan, Steven Vidal Acevedo Colon, Varnika Sinha, Kaan Akşit, and Stefanie Mueller. 2021. SensiCut: Material-Aware Laser Cutting Using Speckle Sensing and Deep Learning. In *The 34th Annual ACM Symposium on User Interface Software and Technology* (Virtual Event, USA) (*UIST '21*). ACM, New York, NY, USA, 24–38. <https://doi.org/10.1145/3472749.3474733>
- [14] Sheila A Dugan and Krishna P Bhat. 2005. Biomechanics and analysis of running gait. *Physical Medicine and Rehabilitation Clinics* 16, 3 (2005), 603–621.
- [15] RV Feethery Jr. 1986. The biomechanics of running on different surfaces. *Clinics in podiatric medicine and surgery* 3, 4 (1986), 649–659.
- [16] A.F. Fercher and J.D. Briers. 1981. Flow visualization by means of single-exposure speckle photography. *Optics Communications* 37, 5 (1981), 326–330. [https://doi.org/10.1016/0030-4018\(81\)90428-4](https://doi.org/10.1016/0030-4018(81)90428-4)
- [17] Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T. Roweis, and William T. Freeman. 2006. Removing Camera Shake from a Single Photograph. *ACM Trans. Graph.* 25, 3 (jul 2006), 787–794. <https://doi.org/10.1145/1141911.1141956>
- [18] Nicholas A. Giudice, Hari Prasath Palani, Eric Brenner, and Kevin M. Kramer. 2012. Learning Non-Visual Graphical Information Using a Touch-Based Vibro-Audio Interface. In *Proceedings of the 14th International ACM SIGACCESS Conference on*

Computers and Accessibility (Boulder, Colorado, USA) (*ASSETS '12*). ACM, New York, NY, USA, 103–110. <https://doi.org/10.1145/2384916.2384935>

[19] Joseph W Goodman. 2007. *Speckle phenomena in optics: theory and applications*. Roberts and Company Publishers, Greenwood Village, CO, USA.

[20] Chris Harrison and Scott E. Hudson. 2008. Lightweight Material Detection for Placement-Aware Mobile Computing. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology* (Monterey, CA, USA) (*UIST '08*). ACM, New York, NY, USA, 279–282. <https://doi.org/10.1145/1449715.1449761>

[21] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, Las Vegas, NV, USA, 770–778.

[22] Wilko Heuten, Daniel Wichmann, and Susanne Boll. 2006. Interactive 3D Sonification for the Exploration of City Maps. In *Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles* (Oslo, Norway) (*NordiCHI '06*). ACM, New York, NY, USA, 155–164. <https://doi.org/10.1145/1182475.1182492>

[23] Youlian Hong, Lin Wang, Jing Xian Li, and Ji He Zhou. 2012. Comparison of plantar loads during treadmill and overground running. *Journal of Science and Medicine in Sport* 15, 6 (2012), 554–560.

[24] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. 2019. Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision*. IEEE, Seoul, Korea (South), 1314–1324.

[25] R Dan Jacobson. 1998. Navigating maps with little or no sight: An audio-tactile approach. In *Content Visualization and Intermedia Representations (CVIR'98)*. CIRSE, Santa Barbara, CA, USA, 95–102.

[26] Ahmad Jalal, Shaharyar Kamal, and Daijin Kim. 2014. A depth video sensor-based life-logging human activity recognition system for elderly care in smart indoor environments. *Sensors* 14, 7 (2014), 11735–11759.

[27] Shaun K. Kane, Meredith Ringel Morris, Annuska Z. Perkins, Daniel Wigdor, Richard E. Ladner, and Jacob O. Wobbrock. 2011. Access Overlays: Improving Non-Visual Access to Large Touch Screens for Blind Users. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (Santa Barbara, California, USA) (*UIST '11*). ACM, New York, NY, USA, 273–282. <https://doi.org/10.1145/2047196.2047232>

[28] Jutta Katharina Willamowski, Shreepriya Gonzalez-Jimenez, Christophe Legras, and Danilo Gallo. 2022. FlexNav: Flexible Navigation and Exploration through Connected Runnable Zones. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (*CHI '22*). ACM, New York, NY, USA, 17 pages. <https://doi.org/10.1145/3491102.3502051>

[29] Young-Ho Kim, Diana Chou, Bongshin Lee, Margaret Danilovich, Amanda Lazar, David E. Conroy, Hernisa Kacorri, and Eun Kyong Choe. 2022. MyMove: Facilitating Older Adults to Collect In-Situ Activity Labels on a Smartwatch with Speech. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (*CHI '22*). ACM, New York, NY, USA, 21 pages. <https://doi.org/10.1145/3491102.3517457>

[30] Yoshiyuki Kobayashi, Takamichi Takashima, Mieko Hayashi, and Hiroshi Fujimoto. 2005. Gait analysis of people walking on tactile ground surface indicators. *IEEE Transactions on neural systems and rehabilitation engineering* 13, 1 (2005), 53–59.

[31] Denys J. C. Matthies, Thijs Roumen, Arjan Kuijper, and Bodo Urban. 2017. CapSoles: Who is Walking on What Kind of Floor?. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vienna, Austria) (*MobileHCI '17*). ACM, New York, NY, USA, 14 pages. <https://doi.org/10.1145/3098279.3098545>

[32] Thomas A McMahon and Peter R Greene. 1979. The influence of track compliance on running. *Journal of biomechanics* 12, 12 (1979), 893–904.

[33] Heiko Müller, Emma Napari, Lauri Hakala, Ashley Colley, and Jonna Häkkilä. 2019. Running Shoe with Integrated Electrochromic Displays. In *Proceedings of the 8th ACM International Symposium on Pervasive Displays* (Palermo, Italy) (*PerDis '19*). ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3321335.3329686>

[34] Elizabeth L. Murnane, Xin Jiang, Anna Kong, Michelle Park, Weili Shi, Connor Soohoo, Luke Vink, Iri Xia, Xin Yu, John Yang-Sammataro, Grace Young, Jenny Zhi, Paula Moya, and James A. Landay. 2020. Designing Ambient Narrative-Based Interfaces to Reflect and Motivate Physical Activity. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (*CHI '20*). ACM, New York, NY, USA, 1–14. <https://doi.org/10.1145/3313831.3376478>

[35] Andrew J. Newell, Ruth M. Morgan, Lewis D. Griffin, Peter A. Bull, John R. Marshall, and Giles Graham. 2012. Automated Texture Recognition of Quartz Sand Grains for Forensic Applications*. *Journal of Forensic Sciences* 57, 5 (2012), 1285–1289. <https://doi.org/10.1111/j.1556-4029.2012.02126.x>

[36] Alex Olwal, Andrew Bardagjy, Jan Zizka, and Ramesh Raskar. 2012. SpeckleEye: Gestural Interaction for Embedded Electronics in Ubiquitous Computing. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems* (Austin, Texas, USA) (*CHI EA '12*). ACM, New York, NY, USA, 2237–2242. <https://doi.org/10.1145/2212776.2223782>

[37] Martin J-D Otis and Bob-Antoine J Menelas. 2012. Toward an augmented shoe for preventing falls related to physical conditions of the soil. In *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, Seoul, Korea (South), 3281–3285.

[38] Yaokun Pang, Xianchen Xu, Shoue Chen, Yuhui Fang, Xiaodong Shi, Yiming Deng, Zhong-Lin Wang, and Changyong Cao. 2022. Skin-inspired textile-based tactile sensors enable multifunctional sensing of wearables and soft robots. *Nano Energy* 96 (2022), 107137. <https://doi.org/10.1016/j.nanoen.2022.107137>

[39] Peter Parente and Gary Bishop. 2003. BATS: the blind audio tactile mapping system. In *Proceedings of the ACM Southeast regional conference*. ACM-SE, New York, NY, USA, 132–137.

[40] Siyou Pei, Pradyumna Chari, Xue Wang, Xiaoying Yang, Achuta Kadambi, and Yang Zhang. 2022. Forcesight: Non-Contact Force Sensing with Laser Speckle Imaging. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) (*UIST '22*). Association for Computing Machinery, New York, NY, USA, Article 25, 11 pages. <https://doi.org/10.1145/3526113.3545622>

[41] Walter Pirker and Regina Katzenschlager. 2017. Gait disorders in adults and the elderly. *Wiener Klinische Wochenschrift* 129, 3 (2017), 81–95.

[42] Giulio Reina and Annalisa Millella. 2012. Towards autonomous agriculture: Automatic ground detection using trinocular stereovision. *Sensors* 12, 9 (2012), 12405–12423.

[43] Prasanta Sahoo and Tapan Kr. Barman. 2012. 5 - ANN modelling of fractal dimension in machining. In *Mechatronics and Manufacturing Engineering*, J. Paulo Davim (Ed.), Woodhead Publishing, Cambridge, United Kingdom, 159–226. <https://doi.org/10.1533/9780857095893.159>

[44] Jaime Sánchez, Mauricio Sáenz, Alvaro Pascual-Leone, and Lotfi Merabet. 2010. Navigation for the Blind through Audio-Based Virtual Environments. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems* (Atlanta, Georgia, USA) (*CHI EA '10*). ACM, New York, NY, USA, 3409–3414. <https://doi.org/10.1145/1753846.1753993>

[45] Munehiko Sato, Shigeo Yoshida, Alex Olwal, Boxin Shi, Atsushi Hiyama, Tomohiro Tanikawa, Michitaka Hirose, and Ramesh Raskar. 2015. SpecTrans: Versatile Material Classification for Interaction with Textureless, Specular and Transparent Surfaces. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 2191–2200. <https://doi.org/10.1145/2702123.2702169>

[46] Shreepriya Shreepriya, Jutta Willamowski, and Danilo Gallo. 2019. Supporting Natural Navigation for Running in Unknown Places. In *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion* (San Diego, CA, USA) (*DIS '19 Companion*). ACM, New York, NY, USA, 277–281. <https://doi.org/10.1145/3301019.3323895>

[47] Karen Simonyan and Andrew Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv:1409.1556 [cs.CV]*

[48] Brandon M Smith, Pratham Desai, Vishal Agarwal, and Mohit Gupta. 2017. CoLux: Multi-object 3d micro-motion analysis using speckle imaging. *ACM Transactions on Graphics (TOG)* 36, 4 (2017), 1–12.

[49] Aiguo Song, Yezhen Han, Haihua Hu, and Jianqing Li. 2014. A Novel Texture Sensor for Fabric Texture Measurement and Classification. *IEEE Transactions on Instrumentation and Measurement* 63, 7 (2014), 1739–1747. <https://doi.org/10.1109/TIM.2013.2293812>

[50] S Strada, A Ghezzi, L Marasco, E Paracampo, G Rizzetto, Patrizia Casali, and Sergiu M Savarese. 2020. Leveraging walking inertial pattern for terrain classification. In *2020 International Conferences on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics (Cybermatics)*. IEEE, Rhodes, Greece, 864–869.

[51] Stephanie Studenski, Subashan Perera, Kushang Patel, Caterina Rosano, Kimberly Faulkner, Marco Inzitari, Jennifer Brach, Julie Chandler, Peggy Cawthon, Elizabeth Barrett Connor, et al. 2011. Gait speed and survival in older adults. *Jama* 305, 1 (2011), 50–58.

[52] Jing Su, Alyssa Rosenzweig, Ashvin Goel, Eyal de Lara, and Khai N. Truong. 2010. Timbremap: Enabling the Visually-Impaired to Use Maps on Touch-Enabled Devices. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services* (Lisbon, Portugal) (*MobileHCI '10*). ACM, New York, NY, USA, 17–26. <https://doi.org/10.1145/1851600.1851606>

[53] Shuochen Su, Felix Heide, Robin Swanson, Jonathan Klein, Clara Callenberg, Matthias Hullin, and Wolfgang Heidrich. 2016. Material classification using raw time-of-flight measurements. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, Las Vegas, NV, USA, 3503–3511.

[54] Wei Sun, Tuochao Chen, Jiayi Zheng, Zhenyu Lei, Lucy Wang, Benjamin Steeper, Peng He, Matthew Dressa, Feng Tian, and Cheng Zhang. 2020. Vibrosense: Recognizing home activities by deep learning subtle vibrations on an interior surface of a house from a single point using laser doppler vibrometry. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–28.

[55] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, Boston, MA, USA, 1–9.

[56] Michael Uelschen and Heinz-Josef Eikerling. 2015. A Mobile Sensor System for Gait Analysis Supporting the Assessment of Rehabilitation Measures. In *Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics* (Atlanta, Georgia) (BCB '15). ACM, New York, NY, USA, 96–105. <https://doi.org/10.1145/2808719.2808729>

[57] Yancheng Wang, Jianing Chen, and Deqing Mei. 2020. Recognition of surface texture with wearable tactile sensor array: A pilot Study. *Sensors and Actuators A: Physical* 307 (2020), 111972. <https://doi.org/10.1016/j.sna.2020.111972>

[58] Jeff Wilson, Bruce N. Walker, Jeffrey Lindsay, Craig Cambias, and Frank Dellaert. 2007. SWAN: System for Wearable Audio Navigation. In *Proceedings of the 2007 11th IEEE International Symposium on Wearable Computers (ISWC '07)*. IEEE Computer Society, USA, 1–8. <https://doi.org/10.1109/ISWC.2007.4373786>

[59] Paweł W. Woźniak, Monika Zbytniewska, Francisco Kiss, and Jasmin Niess. 2021. Making Sense of Complex Running Metrics Using a Modified Running Shoe. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3411764.3445506>

[60] Youcan Yan, Zhe Hu, Yajing Shen, and Jia Pan. 2022. Surface Texture Recognition by Deep Learning-Enhanced Tactile Sensing. *Advanced Intelligent Systems* 4, 1 (2022), 2100076.

[61] Zihan Yan, Yufei Wu, Yang Zhang, and Xiang 'Anthony' Chen. 2022. EmoGlass: An End-to-End AI-Enabled Wearable Platform for Enhancing Self-Awareness of Emotional Health. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) (CHI '22). ACM, New York, NY, USA, 19 pages. <https://doi.org/10.1145/3491102.3501925>

[62] Zihan Yan, Jiayi Zhou, Yufei Wu, Guanhong Liu, Danli Luo, Zihong Zhou, Haipeng Mi, Lingyun Sun, Xiang 'Anthony' Chen, Ye Tao, Yang Zhang, and Guanyun Wang. 2022. Shoes++: A Smart Detachable Sole for Social Foot-to-Foot Interaction. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 2 (jul 2022), 29 pages. <https://doi.org/10.1145/3534620>

[63] Rayoung Yang, Sangmi Park, Sonali R. Mishra, Zhenan Hong, Clint Newsom, Hyeon Joo, Erik Hofer, and Mark W. Newman. 2011. Supporting Spatial Awareness and Independent Wayfinding for Pedestrians with Visual Impairments. In *The Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility* (Dundee, Scotland, UK) (ASSETS '11). ACM, New York, NY, USA, 27–34. <https://doi.org/10.1145/2049536.2049544>

[64] Koji Yatani, Nikola Banovic, and Khai Truong. 2012. SpaceSense: Representing Geographical Information to Visually Impaired People Using Spatial Tactile Feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) (CHI '12). ACM, New York, NY, USA, 415–424. <https://doi.org/10.1145/2207676.2207734>

[65] Hui-Shyong Yeo, Juyoung Lee, Andrea Bianchi, David Harris-Birtill, and Aaron Quigley. 2017. SpeCam: Sensing Surface Color and Material with the Front-Facing Camera of a Mobile Device. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vienna, Austria) (MobileHCI '17). ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3098279.3098541>

[66] Joo Chuan Yeo, Zhuangjian Liu, Zhi-Qian Zhang, Pan Zhang, Zhiping Wang, and Chwee Teck Lim. 2017. Wearable Mechanotransduced Tactile Sensor for Haptic Perception. *Advanced Materials Technologies* 2, 6 (2017), 1700006. <https://doi.org/10.1002/admt.201700006>

[67] Tuo Yu and Klara Nahrstedt. 2019. ShoesHacker: Indoor Corridor Map and User Location Leakage through Force Sensors in Smart Shoes. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3 (sep 2019), 29 pages. <https://doi.org/10.1145/3351278>

[68] Ruidong Zhang, Mingyang Chen, Benjamin Steeper, Yaxuan Li, Zihan Yan, Yizhuo Chen, Songyun Tao, Tuochao Chen, Hyunchul Lim, and Cheng Zhang. 2021. SpeeChin: A Smart Necklace for Silent Speech Recognition. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 4 (2021), 1–23.

[69] Shan Zhang, Zihan Yan, Shardul Sapkota, Shengdong Zhao, and Wei Tsang Ooi. 2021. Moment-to-Moment Continuous Attention Fluctuation Monitoring through Consumer-Grade EEG Device. *Sensors* 21, 10 (2021), 3419. <https://doi.org/10.3390/s21103419>

[70] Yang Zhang, Yasha Iravantchi, Haojian Jin, Swarun Kumar, and Chris Harrison. 2019. Sozu: Self-Powered Radio Tags for Building-Scale Activity Sensing. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) (UIST '19). ACM, New York, NY, USA, 973–985. <https://doi.org/10.1145/3332165.3347952>

[71] Yang Zhang, Gierad Laput, and Chris Harrison. 2018. Vibrosight: Long-Range Vibrometry for Smart Environment Sensing. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (Berlin, Germany) (UIST '18). ACM, New York, NY, USA, 225–236. <https://doi.org/10.1145/3242587.3242608>

[72] Yang Zhang, Gierad Laput, and Chris Harrison. 2018. Vibrosight: Long-Range Vibrometry for Smart Environment Sensing. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (Berlin, Germany) (UIST '18). ACM, New York, NY, USA, 225–236. <https://doi.org/10.1145/3242587.3242608>

[73] Jin Zhou and Baoxin Li. 2006. Homography-based ground detection for a mobile robot platform using a single camera. In *Proceedings 2006 IEEE International Conference on Robotics and Automation, 2006. ICRA 2006*. IEEE, Orlando, FL, USA, 4100–4105. <https://doi.org/10.1109/ROBOT.2006.1642332>

[74] Jan Zizka, Alex Olwal, and Ramesh Raskar. 2011. SpeckleSense: Fast, Precise, Low-Cost and Compact Motion Sensing Using Laser Speckle. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (Santa Barbara, California, USA) (UIST '11). ACM, New York, NY, USA, 489–498. <https://doi.org/10.1145/2047196.2047261>

[75] Markus Zrenner, Christoph Feldner, Ulf Jensen, Nils Roth, Robert Richer, and Björn M. Eskofier. 2020. Evaluation of Foot Kinematics During Endurance Running on Different Surfaces in Real-World Environments. In *Proceedings of the 12th International Symposium on Computer Science in Sport (IACSS 2019)*, Martin Lames, Alexander Danilov, Egor Timme, and Yuri Vassilevski (Eds.). Springer International Publishing, Cham, 106–113.

A CONFIGURATION EXPERIMENT

Table 1: Classification results of different wavelength-and-distance combinations. Accuracy greater than 98% are bolded.

| Wavelength (nm) | Distance (cm) | Accuracy (%) | Wavelength (nm) | Distance (cm) | Accuracy (%) |
|-----------------|---------------|-------------------------|-----------------|---------------|--------------------------|
| 650 (red) | 1 | 82.35 \pm 2.15 | 520 (green) | 1 | 98.09 \pm 1.53 |
| | 3 | 96.84 \pm 1.97 | | 3 | 99.96 \pm 0.23 |
| | 5 | 96.69 \pm 0.97 | | 5 | 98.61 \pm 1.29 |
| | 7 | 99.91 \pm 0.34 | | 7 | 99.60 \pm 0.67 |
| | 9 | 97.49 \pm 1.55 | | 9 | 99.36 \pm 0.79 |
| | 11 | 94.13 \pm 2.14 | | 11 | 99.92 \pm 0.32 |
| | 13 | 98.88 \pm 1.19 | | 13 | 89.09 \pm 0.55 |
| | 15 | 95.44 \pm 1.84 | | 15 | 92.33 \pm 1.72 |
| Wavelength (nm) | Distance (cm) | Accuracy (%) | Wavelength (nm) | Distance (cm) | Accuracy (%) |
| 450 (blue) | 1 | 96.60 \pm 1.92 | 405 (purple) | 1 | 90.80 \pm 2.51 |
| | 3 | 67.48 \pm 2.44 | | 3 | 91.55 \pm 1.82 |
| | 5 | 91.92 \pm 2.16 | | 5 | 99.99 \pm 0.13 |
| | 7 | 99.69 \pm 0.59 | | 7 | 99.53 \pm 0.85 |
| | 9 | 97.76 \pm 1.23 | | 9 | 100.00 \pm 0.00 |
| | 11 | 94.56 \pm 2.49 | | 11 | 91.53 \pm 2.48 |
| | 13 | 98.20 \pm 1.34 | | 13 | 91.91 \pm 0.34 |
| | 15 | 83.69 \pm 0.96 | | 15 | 100.00 \pm 0.00 |