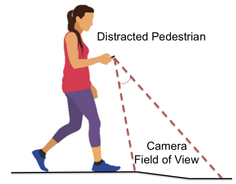
Dataset: Multi-city Street-Sidewalk Imagery from Pedestrian Mobile Cameras

Shubham Jain
Stony Brook University
jain@cs.stonybrook.edu

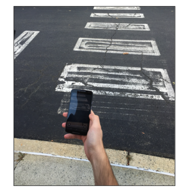
ABSTRACT

This paper presents TerraFirma, a multi-city dataset which captures street and sidewalk imagery from the pedestrians' perspective. Motivated by challenges in the realm of pedestrian safety, we present a diverse and extensive dataset that provides a foundation for the design and validation of pedestrian safety systems that rely on street-sidewalk imagery. The data was collected by 9 volunteers in 4 metropolitan cities across the world. Volunteers carried mobile cameras or smartphones in a texting position, such that the rear camera was directed to the ground in front of them. TerraFirma classifies images by the material used for street/sidewalk construction in each city. The detailed description of dataset accrual is accompanied by the public release of the dataset.

ACM Reference Format:
Shubham Jain. 2020. Dataset: Multi-city Street-Sidewalk Imagery from Pedestrian Mobile Cameras. In *The 3rd International SenSys+BuildSys Workshop on Data: Acquisition to Analysis (DATA '20), November 16–19, 2020, Virtual Event, Japan.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3419016.3431486



(a) Texting pedestrian and smartphone position.



(b) Street-sidewalk frame capture

Figure 1: Data collection process.

1 INTRODUCTION

Texting while walking in public is widely considered a safety risk. Research has attributed a surge in pedestrian fatalities to the phenomenon of distracted walking. Cities have also started printing LOOK! signs at crosswalks in an attempt to warn distracted pedestrian. Smartphones, that are the root cause of distraction, can also be a part of the solution.

Recent efforts in pedestrian safety have focused on using mobile and wearable sensors for detecting pedestrian context [1–6]. Motivated by challenges in the realm of pedestrian safety and Vision Zero [7], we present TerraFirma [8], a diverse and extensive dataset that provides a foundation for the design and validation of pedestrian safety systems that rely on street-sidewalk imagery. The detailed description of dataset accrual is accompanied by the public release of the dataset, as well as initial experience with using the dataset for developing and validating a pedestrian risk detection system [4].

The TerraFirma dataset [8] features images captured using mobile cameras across 4 different cities across the world. The mobile camera was held as shown in Figure 1, such that it captured the walking area (street/sidewalk) in front of the user. A unique contribution of this dataset is its multi-city capture environments. This encompasses the various materials used across the world in the construction of streets and sidewalks.

In addition to detecting pedestrian transitions from sidewalk to street, TerraFirma dataset can be used for other opportunities in pedestrian safety, such as detecting the presence of hazards in the pedestrian's path. Beyond that, this dataset can be of immense use to the computer vision community aiming to build more robust models for material detection and classification. The cities represented in the TerraFirma dataset can immediately monitor the condition of their sidewalks, including information about the crosswalks (visible markings, obstructed crosswalks etc). More importantly, this dataset can form the basis for gathering pedestrian analytics in the city, including but not limited to waiting times at intersections, crossing times, crossing speed variability, crowd estimation etc. Not only is the TerraFirma dataset useful for building systems for pedestrian safety, it can also be helpful in city design and policy. The imagery in the dataset can inform guidelines on material for accommodating disabled population of the city.

2 DATASET DESIGN AND COLLECTION

9 volunteers across 4 metropolitan areas across the world collected data while walking in their cities. Four unique aspects of the data are discussed here. First, it was collected across 4 cities in 3 countries. These cities were New York and Pittsburgh in United States, London in United Kingdom, and Paris in France. Second, it was collected by 9 unique participants, which accounts for the differences in how pedestrians hold their phones when walking. Third, the data was collected from different smartphones, which allows for variability in smartphone cameras. Fourth, our dataset was collected at different times of the year, accounting for seasonal and lighting variations. To summarize, this dataset was collected in-the-wild, in uncontrolled settings.

DATA '20, November 16–19, 2020, Virtual Event, Japan

ACM ISBN 978-1-4503-8136-9/20/11...$15.00
https://doi.org/10.1145/3419016.3431486

(a) Asphalt **(b) Concrete** **(c) Brick** **(d) Tiles** **(e) Ideal Crosswalk** **(f) Crowded**

(g) Rainy **(h) Tiled Pattern** **(i) Deformations** **(j) Defaced Street** **(k) Cluttered** **(l) Same material**

Figure 2: (a)-(d): Samples of material classes found in our dataset. (e)-(l): Examples of collected data.

City	Volunteers	Duration	Sidewalk	Street
New York	5	5h	Concrete	Asphalt
London	1	45m	Tiles Brick	Asphalt Brick
Paris	2	3h 30m	Asphalt	Asphalt Brick
Pittsburgh	1	1h 10m	Concrete Brick	Asphalt Brick

Table 1: TerraFirma Dataset Summary.

Table 1 presents the details of the TerraFirma dataset. The data was collected as real-world videos recorded during the pedestrian's commute and daily chores. As a result, the dataset is biased towards sidewalk imagery, compared to street imagery. To ensure the participants' safety, the videos were captured only during daytime. To ensure privacy, they were asked to record videos only when they were walking alone. However, as a precaution, all the collected videos were stripped of any audio before processing. They were required to hold their phones as they would in a texting position, and use the video recording feature of their phone to record using the rear camera. In a typical texting position, the view of the smartphone's rear camera comprises the ground surface ahead of the user. To avoid any bias, participants were not made aware of the purpose of the data collection. There is a wide variation in the angles that the phone was held in, which adds diversity to our test set.

Data Labeling. The acquired videos were labeled manually. The annotations included frame numbers when the pedestrian set foot into the street to make a sidewalk-street transition, and when they exit the street, to make a street-sidewalk transition. These frame numbers were used as ground truth for demarcating street-sidewalk boundaries. These labels were then used to split each video into two categories - street and sidewalk. The frames that did not capture at least 80% of street or sidewalk were discarded. This was done to ensure that a major part of the frame was occupied by the labelled material, to ensure training fidelity for supervised learning algorithms. For example, frames captured during the transition where street and sidewalk occupied half the frame each, were discarded. However, to maintain the characteristics of an in-the-wild dataset, we retain images with ground artifacts, for example, tree trunks, manhole covers, poles, ramp grates etc, but no crowds. Further, we annotated the frame number in each street/sidewalk video clip where the material of the ground changed. Next, we extracted frames from these clips and manually classified them based on the material they were made of. Frames were extracted at the rate of 3 frames per second. This was done to conserve energy by simulating real-world mobile applications, which is realistic for real-world mobile applications that must operate under tight energy constraints.

Dataset Details. We discuss the details of data collection city-wise. In New York City, the data was collected from the midtown area of Manhattan. This is the same data set as used in LookUp [1]. The camera sensing data was collected by five volunteers, who traversed a 2.1 miles long path. The average time taken to complete each loop was about 60 mins and involved 32 crossings. The data was collected at various times, including weekday rush hours and weekends. This dataset was collected using a GoPro Hero 3 camera. The camera was placed upon the pedestrian, using a chest harness, and oriented to point downwards, simulating the texting position of a smartphone. The GoPro recorded video at 60 frames per second at a resolution of 720p. We downsample this high frame rate data. In London, the data was recorded using a Nexus 6, during daily commutes and weekend chores over a period of two months. In Pittsburgh, the data was recorded in one 70-minute long walking session using an iPhone 6s. It was recorded around dusk, in the downtown area, and also covers two bridges. In Paris, the data was recorded using an LG Nexus 5 during the early morning and afternoon hours, primarily during daily commute.

Figure 2 shows several examples of images from our TerraFirma dataset. The lack of a standard guideline for which materials must be used in paving sidewalks and streets, and frequent changes in

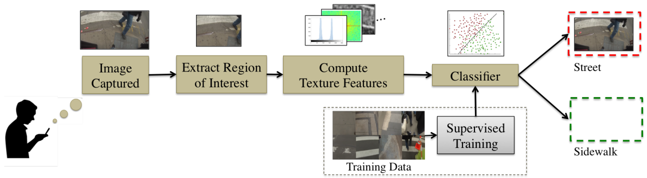


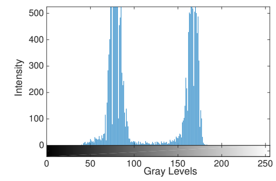
Figure 3: Example use case.

local policies [9] leaves our city sidewalks and streets full of diverse materials. Figures 2(a) - 2(d) show sample images for four materials - asphalt, concrete, brick, and tiles. As may be seen in Figure 2a, some of the captured frames can be blurry because they were extracted from videos while the person was moving. Figure 2e shows an ideal crosswalk with visible stripes. Crosswalks that were made of asphalt have been included under street->asphalt category. Figure 2f shows a sample from the Times Square area in NY where we observed huge crowds. Such samples, where the ground is not visible, were manually removed from the dataset. Figure 2g shows an example from data collected on a rainy day in London. Such samples have been retained in the dataset. Figure 2i shows deformations in the sidewalk, while Figure 2j shows sidewalks painted over. Such examples have also been retained in the dataset. Color of the material is not considered a factor for discarding any images. Only frames where street/sidewalk material occupied less than 80% of the frame, were discarded.

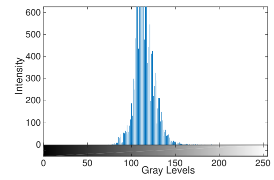
Dataset Structure. The dataset is structured in the following way: the cities form the top level directories (level 1). Inside each city, the next level of directories are street and sidewalk (level 2). Each of these level 2 directories contains multiple directories (level 3), corresponding to the material making up that category (street or sidewalk). For each material, we have included the frames corresponding to that material. For example, the Pittsburgh directory contains 2 directories - street and sidewalk. The street directory further consists of two directories - asphalt and brick. This indicates that for our dataset, the streets captured from Pittsburgh were made of asphalt or brick. In total, there are approximately 6000 images in the TerraFirma dataset.

3 DATA ANALYSIS USING TERRAFIRMA

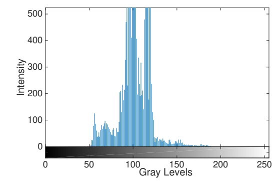
TerraFirma has been used for designing and validating a pedestrian safety application [4]. The high-level pipeline for the system is shown in Figure 3. The system performs pedestrian risk detection by identifying when a user is about to transition from safe walking areas (sidewalk) to less safe walking areas (street). This is done by training a classifier that can distinguish between streets and sidewalk. The Terrafirma dataset was used for understanding texture from images captured using a smartphone. This is significant, since traditionally texture-based analysis has been performed on elaborate setups with specialized cameras. We extract texture-based features from the images. Figure 4 shows sample intensity histograms from data collected in New York City. The number of peaks in the intensity histogram, and the spacing between them can be used to distinguish a cluttered image from one with a pattern, possibly a crosswalk. The system demonstrates that mobile cameras can be enabled to distinguish materials of walking surfaces in urban areas with more than 90% accuracy, and accurately identify when pedestrians transition from sidewalk to street. Such a pedestrian safety system can potentially alert pedestrians or communicate safety information to oncoming vehicles.



(a) Crosswalk



(b) Concrete Sidewalk



(c) Crowded Sidewalk

Figure 4: Intensity Histograms.

4 CONCLUSION

We present TerraFirma, a multi-city dataset of street-sidewalk imagery, captured by pedestrians during their regular commute. The data was collected in 4 metropolitan cities of the world by 9 volunteers. The utility of this dataset has previously been demonstrated in the design and validation of a pedestrian safety system. We hope that this exhaustive dataset and annotations will help researchers train and validate their vision algorithms.

REFERENCES

[1] Shubham Jain, Carlo Borgiattino, Yanzhi Ren, Marco Gruteser, Yingying Chen, and Carla Fabiana Chiasserini. Lookup: Enabling pedestrian safety services via shoe sensing. In *MobiSys ’15*, MobiSys ’15, pages 257–271, New York, NY, USA, 2015. ACM.
[2] Shubham Jain, Carlo Borgiattino, Yanzhi Ren, Marco Gruteser, and Yingying Chen. On the limits of positioning-based pedestrian risk awareness. In *Proceedings of the 2014 Workshop on Mobile Augmented Reality and Robotic Technology-based Systems*, MARS ’14, 2014.
[3] T. Datta, S. Jain, and M. Gruteser. Towards city-scale smartphone sensing of potentially unsafe pedestrian movements. In *HotPlanet 2014*, pages 663–667, Oct 2014.
[4] S. Jain and M. Gruteser. Recognizing textures with mobile cameras for pedestrian safety applications. *IEEE Transactions on Mobile Computing*, 18(8):1911–1923, 2019.
[5] Yu-Chih Tung and Kang G Shin. Use of phone sensors to enhance distracted pedestrians’ safety. *IEEE Transactions on Mobile Computing*, 17(6):1469–1482, 2017.
[6] Maozhi Tang, Cam-Tu Nguyen, Xiaoliang Wang, and Sanglu Lu. An efficient walking safety service for distracted mobile users. In *2016 IEEE 13th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, pages 84–91. IEEE, 2016.
[7] Vision zero. https://www1.nyc.gov/content/visionzero/pages/.
[8] Terrafirma dataset. https://doi.org/10.5061/dryad.2280gb5qj.
[9] Walkable Princeton. Council decision means princeton’s brick crosswalks will disappear. 2013.