SideSeeing: A multimodal dataset and tools for sidewalk assessment

Rafael J. P. Damaceno*, Lenardo Ferreira†, Fabio Miranda†,
Maryam Hosseini‡, Roberto M. Cesar-Jr.*
*University of São Paulo (USP)
†University of Illinois Chicago (UIC)
‡University of California Berkeley (UCB)

***Abstract*—This paper introduces SideSeeing, a novel initiative that provides tools and datasets for assessing the built environment. We present a framework for street-level data acquisition, loading, and analysis. Using the framework, we collected a novel dataset that integrates synchronized video footage captured from chest-mounted mobile devices with sensor data (accelerometer, gyroscope, magnetometer, and GPS). Each data sample represents a path traversed by a user filming sidewalks near hospitals in Brazil and the USA. The dataset encompasses three hours of content covering 12 kilometers around nine hospitals, and includes 325,000 video frames with corresponding sensor data. Additionally, we present a novel 68-element taxonomy specifically created for sidewalk scene identification. SideSeeing is a step towards a suite of tools that urban experts can use to perform in-depth sidewalk accessibility evaluations. SideSeeing data and tools are publicly available at https://sites.usp.br/sideseeing/.**

I. Introduction

In recent years, urban informatics and urban computing have opened new horizons for tackling a number of societal problems integral to urban planning and design (Marasinghe et al., 2023, Miranda et al., 2024, Wang and Biljecki, 2022) such as urban accessibility, risk assessment and mapping, climate change, and heat exposure.

These data can be represented in various formats, for instance, visual data captured by video cameras, sound data captured by microphones, textual data extracted from region-related databases, and temporal data obtained from sensors in mobile devices. Real-world urban informatics scenarios are inherently multimodal, integrating these diverse data types to provide comprehensive insights.

Despite the recent surge in urban-centric research, there remains a gap in the availability of comprehensive datasets describing urban public spaces, particularly those dedicated to pedestrians (Deitz et al., 2021). While some datasets may offer information about sidewalks accessibility (Park et al., 2020, Saha et al., 2019), pedestrian networks (Hosseini et al., 2023), sidewalk surface materials (Hosseini et al., 2022), there is a lack of fine-grained, ground-level data on sidewalk conditions that reflect the lived experiences of their users.

In this context, two initiatives developing frameworks and datasets in urban analytics are *Urban Mosaic* (Miranda et al., 2020) and *Footpath.ai*[1]. However, both initiatives primarily concentrate on street-level image data, which may limit their ability to address the broader complexities of urban environments. Furthermore, integrating multimodal data has the potential to improve our understanding of accessibility in urban areas.

To address this gap, we introduce SideSeeing, a novel initiative based on multimodal data that provides tools and datasets for assessing the built environment. SideSeeing includes a framework for street-level data acquisition, loading, and analysis. It also includes a novel dataset that integrates synchronized video footage captured from chest-mounted mobile devices with sensor data (accelerometer, gyroscope, magnetometer, and GPS). The dataset is focused on environments near hospitals, areas that are of particularly importance for public health and accessibility. As Seetharaman et al. (2024) note, gaining a better understanding of the built environment, particularly outdoors spaces, can provide key information to the design of accessible public spaces. By collecting synchronized data from multiple sensors and sources, including video, GPS, and IMU sensors available on mobile devices, we offer a multimodal dataset for analyzing urban environments and assessing accessibility in different contexts.

The main contributions of this paper are: 1) A multimodal dataset composed of urban scenes representing sidewalks near hospitals, called SideSeeing Hospital Dataset; 2) An open-source Android application for collecting synchronized multimodal data using smartphones; 3) A novel taxonomy specifically designed for characterize urban scenes focused on sidewalks; and 4) A Python library for loading and analyzing datasets created with our framework.

II. Proposed System

SideSeeing follows a workflow (Figure 1) of data collection, preprocessing, visualization, AI model development, and information analysis. A chest-mounted smartphone captures video and sensor data, focusing on ground-level urban scenes.

Fig. 1: Workflow diagram showing the steps of SideSeeing from data acquisition to analysis.

A chest mount holds smartphones during data collection, allowing for adjustable angles. Data is collected with ac-

[1]*Footpath.ai* is a company (whose website is https://footpath.ai) dedicated to creating maps from a pedestrian-centric perspective, focusing on walking infrastructure.

celerometer, gyroscope, magnetometer, and GPS sensors, with 1280x720 or 1920x1080 resolution video at 30 frames per second.

For the data collection, we developed an Android-based application that utilizes Google's sensor and camera frameworks. The open-source application, named "MultiSensor Data Collection", is publicly available[2]. It offers various settings that can be customized to a study's specific needs. For the SideSeeing Hospital Dataset, the data collection protocol involved defining a set of parameters, including walking paths, starting angles configured in the mobile application, and device orientation. Subsequently, we will explain the rationale behind our choices for building the dataset using this protocol.

A. Sidewalk Paths

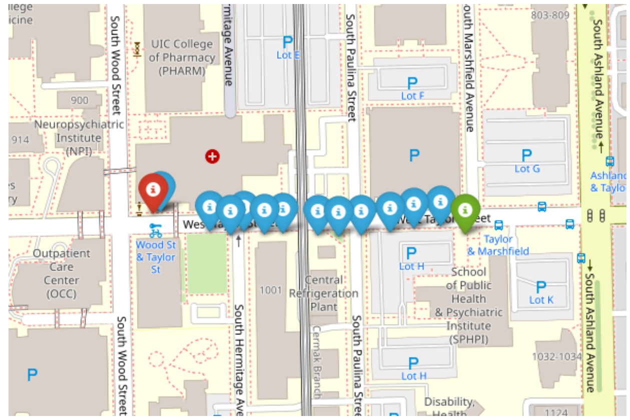
Initially, we mapped out the routes in the four cities where the project participants are located (Santos, Jundiaí, São Paulo, and Chicago). As one of our project goals is to provide information regarding accessibility, we chose to collect data near hospitals, facilities where improper urban access can severely affect people's health. Through the Google Maps platform, we searched for bus stops and train or metro stations near the hospitals we had decided to cover. Each route starts at a public transportation stop and ends at the main entrance of a hospital (or vice-versa). The goal was to simulate a person's route traveling to and from the hospital. Figure 2 illustrates two paths: one showing a path for a person walking from a hospital in Chicago, Illinois, USA to a public transportation stop (Figure 2a), and the other showing a second path for a person walking from a public transportation stop to the hospital's entrance (Figure 2b).

B. Mobile Application Settings

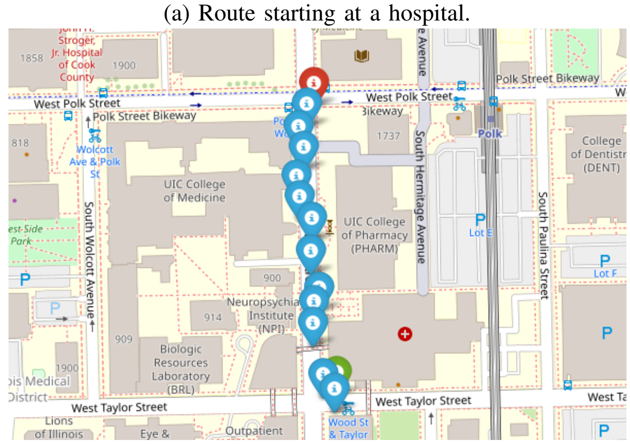
To facilitate adjusting the recording angle, the mobile application presents the angle value to the user before starting the data acquisition. The objective is to maximize the portion of the video showing the sidewalk. This configuration relies on using the default back camera of the device, which must have a wide enough field of view to capture the entire sidewalk width. We aim to utilize only the available cameras on mobile devices, avoiding the need for additional hardware or specific mobile devices (such as fisheye cameras). After a set of initial trials, we opted for an angle of approximately 70 degrees (relative to the ground, where zero degrees would be parallel to the ground, 90 degrees would be perpendicular) to focus on the ground, ensuring optimal sidewalk capture during the recordings.

Another parameter is the device orientation, which can be landscape or portrait. We tested both options, but in the context of the SideSeeing Hospital Dataset, we opted to use the landscape mode. We argue that by using the mobile application in this mode and with an adequate application angle, the video shows a greater level of detail and proximity to where the user is walking. Compounding with the other types of information collected by the mobile application, the potential to enrich the analysis regarding the sidewalk tends to be greater, enabling a more detailed examination of the sidewalk's surface material and conditions.

[2]https://github.com/rafaelpezzuto/multi-sensor-data-collection, accessed on June 10, 2024.



(a) Route starting at a hospital.



(b) Route ending at a hospital.

Fig. 2: Example of routes traced for a person to walk between public transportation stops and a hospital in Chicago, Illinois, USA. The green marker represents the endpoint, and the red marker represents the starting point.

C. Dataset Acquisition

For data acquisition, we wore the chest support, configured the device by activating its auto-rotation and localization options, opened the mobile application, and started recording. The procedure involves pausing for two seconds (approximately), walking along the planned route, and pausing for another two seconds. As shown in Figure 3, these pauses are particularly useful for post-editing steps based on the accelerometer data, such as video splitting. These exact steps were conducted for all the routes.

To further illustrate sensor information from Table I, we present a plot of the gyroscope data in 4a. The yellow highlighted areas on the time series indicate a left turn made by the person recording the data. Similarly to the accelerometer case discussed earlier, this information is valuable for users to divide videos based on pedestrian movements. Figure 4b

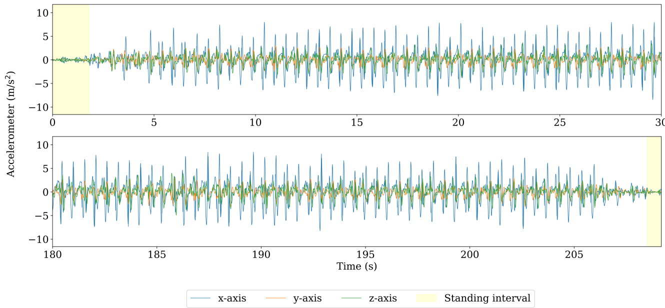
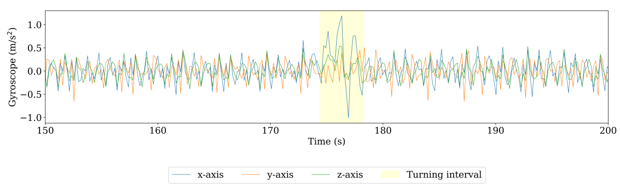
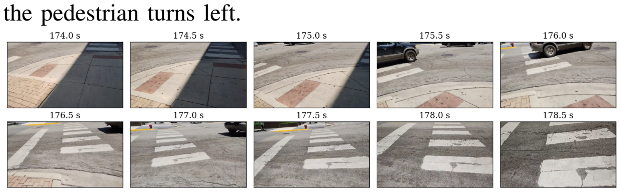


Fig. 3: Accelerometer data revealing pause (yellow regions) and walking periods during a walking path. These data points were extracted from a route recorded in Chicago, Illinois, USA.

shows video frames corresponding to those moments when the pedestrian made turns.



(a) Gyroscope data highlighted in yellow for the moment when the pedestrian turns left.



(b) Video frames for the segment where the pedestrian turns left.

Fig. 4: Gyroscope data revealing turns during a walking path. These data points were extracted from a route recorded in Chicago, Illinois, USA.

D. Dataset Statistics

The dataset includes 46 videos across 12 kilometers, focusing on sidewalk accessibility near hospitals in Brazil and the USA. Table II details the data collected. After video collection, we manually analyzed the footage to identify urban scene elements. Based on this analysis and recent studies (Duan et al., 2021, Gamache et al., 2017, Li and Loftness, 2021), we developed a two-level taxonomy of sidewalk characteristics, including pavement conditions, materials, obstacles, structures, and adjacent road types (Table III). A detailed description of the taxonomy development will be presented in a future manuscript.

E. SideSeeing Python Library

We also developed a Python library, "sideseeing-tools," for data analysis. It synchronizes sensor, video, and GPS data, allowing for visualization and processing via various plotting functions. More details are available at https://pypi.org/project/sideseeing-tools/.

III. Conclusion

SideSeeing provides a comprehensive framework for sidewalk assessment using multimodal data. The dataset, tools, and mobile app enable urban researchers to study sidewalk accessibility. Future work will extend data collection to other cities and explore applications in gait analysis and accessibility classification.

References

Shiloh Deitz, Amy Lobben, and Arielle Alferez. Squeaky wheels: Missing data, disability, and power in the smart city. *Big Data & Society*, 8(2):20539517211047735, 2021. doi:10.1177/20539517211047735.

Michael Duan, Aroosh Kumar, Michael Saugstad, Aileen Zeng, Ilia Savin, and Jon E. Froehlich. Sidewalk Gallery: An interactive, filterable image gallery of over 500,000 sidewalk accessibility problems. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*, pages 1–5, 2021. doi:10.1145/3441852.3476542.

Stephanie Gamache, Yan Grenier, Patrick Fougeyrollas, Geoffrey Edwards, and Mir Abolfazl Mostafavi. Developing a taxonomy of the built environment for disability studies. methodological insights. *Journal of Accessibility and Design for All*, 7(2):236–265, Nov. 2017. doi:10.17411/jacces.v7i2.130.

Maryam Hosseini, Fabio Miranda, Jianzhe Lin, and Claudio T. Silva. CitySurfaces: City-scale semantic segmentation of sidewalk materials. *Sustainable Cities and Society*, 79:103630, 2022. ISSN 2210-6707. doi:10.1016/j.scs.2021.103630.

Maryam Hosseini, Andres Sevtsuk, Fabio Miranda, Roberto M Cesar Jr, and Claudio T Silva. Mapping the walk: A scalable computer vision approach for generating sidewalk network datasets from aerial imagery. *Computers, Environment and Urban Systems*, 101:101950, 2023. doi:10.1016/j.compenvurbsys.2023.101950.

Zekun Li and Vivian Loftness. Mitigating and adapting to climate change with a taxonomy of smart urban surfaces. *Journal of Physics: Conference Series*, 2069(1):012225, nov 2021. doi:10.1088/1742-6596/2069/1/012225.

Raveena Marasinghe, Tan Yigitcanlar, Severine Mayere, Tracy Washington, and Mark Limb. Computer vision applications for urban planning: A systematic review of opportunities and constraints. *Sustainable Cities and Society*, page 105047, 2023. doi:10.1016/j.scs.2023.105047.

Fabio Miranda, Maryam Hosseini, Marcos Lage, Harish Doraiswamy, Graham Dove, and Cláudio T. Silva. Urban Mosaic: Visual exploration of streetscapes using large-scale image data. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–15, New York, NY, USA, 2020. Association for Computing Machinery. ISBN 9781450367080. doi:10.1145/3313831.3376399.

TABLE I: Files created by the MultiSensor Data Collection application for each collected instance.

File	Description
consumption.csv	Data related to power consumption
gps.csv	GPS data
metadata.json	Metadata such as the device settings and application settings
sensors.one.csv	Data from one-axis sensors
sensors.three.csv	Data from three-axis sensors
sensors.three.uncalibrated.csv	Data from three-axis sensors uncalibrated
video.mp4	MP4 video file including audio

TABLE II: Total distance in meters, duration in seconds, number of routes, number of hospitals, total video frames, and total data points for Accelerometer (ACC), gyroscope (GYR), and magnetometer (MAG) across cities.

						Total Data Points		
City - Country	Routes	Hospitals	Distance	Duration	Video Frames	ACC	GYR	MAG
Chicago - USA	25	3	5,865	4,616	175,847	580,758	29,124	303,137
Jundiaí - Brazil	7	2	1,482	1,890	44,476	139,294	8,243	68,096
Santos - Brazil	11	2	2,247	3,532	67,431	215,573	12,499	105,680
São Paulo - Brazil	4	2	1,271	1,840	38,129	92,638	63,624	38,073
All	47	9	10,865	11,878	325,883	1,028,263	113,490	514,986

TABLE III: Taxonomy of sidewalks with two levels. Level one includes adjacent road types, obstacles, pavement conditions, sidewalk geometry, sidewalk structure, and surface materials.

Level 1	Level 2	Level 1	Level 2
Adjacent road type	Motorway / highway	Pavement condition	Broken
	None		Corrugation
	Residential		Cracked
	Service		Detached
Obstacles	Aerial vegetation		Patching
	Barrier		Pothole
	Bench	Sidewalk geometry	Height difference
	Bike rack		Narrow
	Black ice		Steep
	Bus stop	Sidewalk structure	Bioswale
	Car barrier		Curb ramp
	Construction material		Footbridge
	Dirt		Friction strip
	Fence		Grate
	Fire hydrant		Ramp
	Floor standing board		Stairs
	Garage entrance		Tactile paving
	Ground light	Surface type	Asphalt
	Ground vegetation		Bluestone
	Manhole cover		Brick
	Newsstand		Coating
	Parked vehicle		Cobblestone
	Parking booth		Concrete
	Person		Concrete with aggregates
	Pole		Grass
	Potted plant		Gravel
	Puddle		Large pavers
	Rock		Red brick
	Snow		Slab
	Telephone booth		Stone pavement
	Traffic cone		Tiles
	Transit sign		
	Trash can		
	Tree leaves		
	Trunck		
	Water channel		
	Water fountain		

Fabio Miranda, Thomas Ortner, Gustavo Moreira, Maryam Hosseini, Milena Vuckovic, Filip Biljecki, Claudio T. Silva, Marcos Lage, and Nivan Ferreira. The state of the art in visual analytics for 3D urban data. *Computer Graphics Forum*, 43(3):e15112, 2024. doi:10.1111/cgf.15112.

Kibaek Park, Youngtaek Oh, Soomin Ham, Kyungdon Joo, Hyokyoung Kim, Hyoyoung Kum, and In So Kweon. Sideguide:a large-scale sidewalk dataset for guiding impaired people. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 10022–10029, 2020. doi:10.1109/IROS45743.2020.9340734.

Manaswi Saha, Michael Saugstad, Hanuma Teja Maddali, Aileen Zeng, Ryan Holland, Steven Bower, Aditya Dash, Sage Chen, Anthony Li, Kotaro Hara, et al. Project Sidewalk: A web-based crowdsourcing tool for collecting sidewalk accessibility data at scale. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2019. doi:10.1145/3290605.3300292.

Kishore Seetharaman, Atiya Mahmood, Farinaz Rikhtehgaran, Ghazaleh Akbarnejad, Farrukh Chishtie, Mike Prescott, Alison Chung, and W Ben Mortenson. Influence of the built environment on community mobility of people living with visual disabilities: a scoping review. *Urban, Planning and Transport Research*, 12(1):2296891, 2024. doi:10.1080/21650020.2023.2296891.

Jing Wang and Filip Biljecki. Unsupervised machine learning in urban studies: A systematic review of applications. *Cities*, 129:103925, 2022. ISSN 0264-2751. doi:10.1016/j.cities.2022.103925.

BIO

Rafael J. P. Damaceno received his PhD in Computer Science from the Federal University of ABC. He has experience in the fields of Human-Computer Interaction and Scientometrics. Since his master's, he has participated in studies on mobile device accessibility. Currently, he is a postdoctoral researcher at the University of São Paulo, working on Computer Vision and Deep Learning for Urban Informatics.

Leonardo Ferreira is a PhD student at the University of Illinois Chicago. He earned his master's degree in Computer Science from Universidade Federal Fluminense. His research focuses on information management, visualization, and multimodal learning to support information conveyance and urban planning.

Fabio Miranda is an assistant professor at the University of Illinois Chicago. He received his PhD in Computer Science from New York University. His research focuses on developing techniques that allow for the interactive analysis of large-scale data, combining methods from visualization, GIS, data management, machine learning, and computer graphics.

Maryam Hosseini is an incoming assistant professor at the University of California, Berkeley. She was a postdoc associate at MIT and a research associate at NYU Tandon's Computer Science Department. She holds a Ph.D. in Urban Systems from Rutgers University. Her work lies at the intersection of urban planning, economics, and computer science, with a focus on advancing the understanding of built environment quality.

Roberto M. Cesar Jr. is a Full Professor of Computer Science at the University of São Paulo. His research interests include computer vision, pattern recognition, image processing, bioinformatics, and eScience. He has a Ph.D. in Physics and has served as a member of various academic committees and as the director of the eScience Research Center at USP.

RATIONALE

Sidewalk conditions directly impact pedestrian safety and comfort, specially near hospitals, where individuals with diverse health conditions often pass. Well-maintained sidewalks also promote social inclusion and support the sustainable use of urban space. Through SideSeeing, we provide a framework for collecting and analysing sidewalk data using video and sensor from smartphones. This integrated data can empower urban planners with insights needed for informed decision-making.