A Feasibility Study of Using Google Street View and Computer Vision to Track the Evolution of Urban Accessibility

Ladan Najafizadeh
University of Maryland, College Park
ladan.n@gmail.com

Jon E. Froehlich
University of Washington
jonf@cs.washington.edu

Figure 1. In this paper, we examine the feasibility of using Google Street View's "time machine" feature [4] and basic computer vision algorithms to track changes in urban accessibility over time. For each location, accessibility problems are manually labeled in the most recent Street View image (blue outline) then are automatically back propagated through time (red outlines) to track and discover potential changes. In the example here, an object in the pedestrian path has persisted over time to the most recent data (2014), while a sidewalk surface problem from 2007 was resolved by 2009.

ABSTRACT
Previous work has explored scalable methods to collect data on the accessibility of the built environment by combining manual labeling, computer vision, and online map imagery. In this poster paper, we explore how to extend these methods to track the evolution of urban accessibility over time. Using Google Street View's "time machine" feature, we introduce a three-stage classification framework: (i) manually labeling accessibility problems in one time period; (ii) classifying the labeled image patch into one of five accessibility categories; (iii) localizing the patch in all previous snapshots. Our preliminary results analyzing 1633 Street View images across 376 locations demonstrate feasibility.

Author Keywords
Urban accessibility; computer vision; Google Street View

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI)

INTRODUCTION
Recent work has explored scalable methods to identify and characterize accessibility features in the built environment using remote crowdsourcing, machine learning, and online map datasets (*e.g., Google Street View* (GSV) [5, 7, 11], satellite photographs [1]). For example, *Tohme* [7] combines computer vision with web-based crowd work to semi-automatically label curb ramps in GSV. While accurately finding and assessing accessibility features in map imagery is still an active research area, in this poster paper, we begin to explore a related but even more data-intensive process—how to semi-automatically track the evolution of urban accessibility over time using historical map data (Figure 1).

Our work builds on decades of past research in urban studies, geography, and ecology, which analyze temporal changes in

land use from remote sensors. Typically, however, the focus is on macroscopic trends (*e.g.,* urbanization [8, 14, 18], deforestation [13]), which do not require the detailed sensing of small entities that our work requires (*e.g.,* light poles, curb ramps). In addition, rather than rely on satellite images, we use the historical omnidirectional panoramic imagery found in GSV's "time machine" [4]. With the emergence of large-scale image sets and an interest in vision algorithms to support autonomous vehicles, computer scientists have also begun to develop techniques to detect and model urban change [2, 9, 12]. Our techniques are informed by these approaches but with a distinct focus on tracking accessibility.

Our contributions include: (i) a preliminary examination of using GSV's "time machine" as a data source for tracking (in)accessible pedestrian infrastructure over time; (ii) an initial three-stage classification framework for labeling and categorizing accessibility features through time; (iii) a preliminary study validating our approach.

FEASIBILITY STUDY
To examine the feasibility of our approach, we created a test dataset, implemented a classification framework, and performed initial validation. Based on [6, 11], we track five classes of sidewalk features: *accessible sidewalks* (*i.e.*, no problems), *accessible curb ramps*, *missing curb ramps*, *objects in path*, and *surface problems.*

Dataset
We built our dataset by randomly selecting locations in Washington DC and Maryland, examining the GSV imagery to identify accessibility features, and then using "time machine" to capture historical panoramas. As we are primarily interested in how accessibility features change over time, we iteratively diversified the dataset to include locations where features: (i) changed over time; (ii) persisted over time; or (iii) were occluded in at least one time period (*e.g.,* by a passing car), making it difficult to track temporal changes. For each location, we captured a screenshot of all available images across time and recorded GPS coordinates, Street View URL, capture timestamp, and the camera's yaw, pitch, and field-of-view.

In total, the dataset includes 1633 images at 376 locations (332 in DC; 44 in Maryland), 90% of which contain at least one accessibility problem; see Table 1. While Google does not publish how often they update GSV imagery, the dataset includes 4.3 snapshots on average per location (SD=1; range: 2–9). The dataset was collected in 2015, with image timestamps roughly every two years back to 2007. As snapshots can occur at slightly different camera positions due to differences in the Google car position, we manually aligned the images at each location—a process that will need to be automated in the future.

	Missing ramps	Objects in path	Surface problems	Accessible Sidewalks	Curb ramp	Total
Per Location	52	96	123	35	70	376
Per Image	231	449	374	285	267	1606

Table 1. Our study dataset broken down by category. *Per Location* refers to the existence of an accessibility feature in at least one snapshot; *Per Image* represents category counts across all images (regardless of location or time).

Approach
Our initial approach employs a three-stage framework: manually outlining an accessibility feature in a single time period, categorizing the outlined feature, then back propagating that label through time. For this feasibility study, we used Matlab's built-in *Image Labeler* and *Computer Vision Toolbox System*.

Stage 1: Manual Labeling. For the most recent image at each location, we manually draw bounding boxes around identified accessibility features (similar to [6]). Example image patches are shown in Figure 2.

Stage 2: Categorizing Image Patches. To categorize the image patches, we use a *Bag of Visual Words* (BoVW) approach [3]—a common image classification technique

Figure 2. Example image patches extracted in Stage 1.

inspired by natural language processing. BoVW works by extracting local descriptors that are repeatable and invariant to image transformations (*e.g.,* translation, scaling) and building a "visual vocabulary" by clustering features into visual words (analogous to words in text documents). A supervised machine learning model—in our case, a multi-class SVM—can then be trained to classify feature vectors into categories.

Stage 3: Detecting Objects through Time. While BoVW provides a robust approach to categorize pre-segmented image patches, it cannot be used to localize accessibility features in previous snapshots. For this, we use Matlab's built-in implementation of the *Viola-Jones* cascade object detector [17] with *Histogram of Oriented Gradients* (HOG) input features—similar to Tohme [7]. For training, we provide a large set of negative examples (image patches that do not contain accessibility features) and positive examples

	Missing ramp	Objects in path	Surface problem	Access. sidewalk	Curb ramp
Missing ramp	66.3%	1.6%	0.1%		12.6%
Objects in path		97.3%			0.6%
Surface problem	12.8%		81.4%	14.1%	3.3%
Accessible sidewalk	11.7%		17.2%	85.7%	9.3%
Curb ramp	10.2%	1.1%	0.3%	0.2%	74.2%

Table 2. A confusion matrix of Stage 2 performance.

(image patches with accessibility features). While the number of classifier stages is a function of dataset size, we found that 12-15 stages performed best. We trained five object detectors for the five accessibility categories. To improve accuracy and speed, we reduce the search area of the sliding window to within 300 pixels of the original Stage 1 bounding box (as the *x,y* pixel location does not change significantly across time given our manual alignment).

Preliminary Experiments and Results
We first evaluate the performance of the category classifier (Stage 2) before analyzing the full three-stage framework.

Category Classification. To evaluate the performance of our image patch classifier alone, we used k-fold cross validation ($k=5$; each fold consisted of 326 image patches). Overall, our classification performance varied from 66.3% for *missing curb ramps* to 97.3% for *objects in path*, which is consistent with prior work [7]. See Table 2.

System Evaluation. To evaluate the performance of our classifier and object detector combined, we randomly split our dataset (per location) into 70% training ($N=264$) and 30% test ($N=112$). The input for each location is the bounding box from Stage 1. To measure correctness, we examined whether the image patch was correctly categorized and found in previous images at each location. Overall, our approach was able to find 77% of the labeled accessibility features (recall) in historical images with a precision of 78%. Missing curb ramps again fared worse: 63% recall and 58% precision most likely because they lack strong discriminant visual features. Due to the manual alignment of historical images, these results should be considered preliminary.

DISCUSSION AND CONCLUSION
The long-term goal of this work is to develop accurate and scalable methods for tracking how urban accessibility changes over time. With the resulting temporal datasets, our hope is to enable new data-driven research inquiries such as: what geo-spatial and socio-economic factors seem to correlate with changes in urban accessibility? How do temporal patterns of urban accessibility compare across cities in the US and beyond? What are the appropriate models and metrics to capture this change? Can we build predictive models of accessibility evolution? While our early results demonstrate feasibility, we plan to build a larger test dataset and conduct a more comprehensive evaluation, integrate state-of-the-art deep learning methods for object identification, and experiment with methods to reduce manual labeling requirements.

REFERENCES

[1] Ahmetovic, D., Manduchi, R., Coughlan, J.M. and Mascetti, S. 2015. Zebra Crossing Spotter: Automatic Population of Spatial Databases for Increased Safety of Blind Travelers. *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility - ASSETS '15* (New York, New York, USA, 2015), 251–258.

[2] Alcantarilla, P.F., Stent, S., Ros, G., Arroyo, R. and Gherardi, R. 2018. Street-view change detection with deconvolutional networks. *Autonomous Robots*. (May 2018), 1–22. DOI:https://doi.org/10.1007/s10514-018-9734-5.

[3] Csurka, G., Dance, C.R., Fan, L., Willamowski, J. and Bray, C. 2004. Visual categorization with bags of keypoints. *In Workshop on Statistical Learning in Computer Vision at ECCV 2004* (2004), 1--22.

[4] Go back in time with Street View: 2014. *https://googleblog.blogspot.com/2014/04/go-back-in-time-with-street-view.html*.

[5] Hara, K. and Froehlich, J.E. 2015. Characterizing and visualizing physical world accessibility at scale using crowdsourcing, computer vision, and machine learning. *ACM SIGACCESS Accessibility and Computing*. 113 (Nov. 2015), 13–21. DOI:https://doi.org/10.1145/2850440.2850442.

[6] Hara, K., Le, V. and Froehlich, J. 2013. Combining Crowdsourcing and Google Street View to Identify Street-level Accessibility Problems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, May 2013), 631–640.

[7] Hara, K., Sun, J., Moore, R., Jacobs, D. and Froehlich, J. 2014. Tohme: detecting curb ramps in google street view using crowdsourcing, computer vision, and machine learning. *Proceedings of the 27th annual ACM symposium on User interface software and technology - UIST '14* (New York, New York, USA, 2014), 189–204.

[8] Ji, W., Ma, J., Twibell, R.W. and Underhill, K. 2006. Characterizing urban sprawl using multi-stage remote sensing images and landscape metrics. *Computers, Environment and Urban Systems*. 30, 6 (Nov. 2006), 861–879. DOI:https://doi.org/10.1016/J.COMPENVURBSYS.2005.09.002.

[9] Matzen, K. and Snavely, N. 2014. Scene Chronology. Springer, Cham. 615–630.

[10] Qin, R., Tian, J. and Reinartz, P. 2016. 3D change detection – Approaches and applications. *ISPRS Journal of Photogrammetry and Remote Sensing*. 122, (Dec. 2016), 41–56. DOI:https://doi.org/10.1016/J.ISPRSJPRS.2016.09.013.

[11] Saha, M., Hara, K., Behnezhad, S., Li, A., Saugstad, M., Maddali, H., Chen, S. and Froehlich, J.E. 2017. A Pilot Deployment of an Online Tool for Large-Scale Virtual Auditing of Urban Accessibility. *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '17* (New York, New York, USA, 2017), 305–306.

[12] Schindler, G. and Dellaert, F. 2010. Probabilistic temporal inference on reconstructed 3D scenes. *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Jun. 2010), 1410–1417.

[13] Skole, D. and Tucker, C. 1993. Tropical deforestation and habitat fragmentation in the Amazon: satellite data from 1978 to 1988. *Science (New York, N.Y.)*. 260, 5116 (Jun. 1993), 1905–10. DOI:https://doi.org/10.1126/science.260.5116.1905.

[14] Song, X.-P., Sexton, J.O., Huang, C., Channan, S. and Townshend, J.R. 2016. Characterizing the magnitude, timing and duration of urban growth from time series of Landsat-based estimates of impervious cover. *Remote Sensing of Environment*. 175, (Mar. 2016), 1–13.

[15] Taneja, A., Ballan, L. and Pollefeys, M. 2013. City-Scale Change Detection in Cadastral 3D Models Using Images. *2013 IEEE Conference on Computer Vision and Pattern Recognition* (Jun. 2013), 113–120.

[16] Taneja, A., Ballan, L. and Pollefeys, M. 2011. Image based detection of geometric changes in urban environments. *2011 International Conference on Computer Vision* (Nov. 2011), 2336–2343.

[17] Viola, P. and Jones, M. 2001. Rapid Object Detection using a Boosted Cascade of Simple Features. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*. 1, (2001), 511. DOI:https://doi.org/http://doi.ieeecomputersociety.org/10.1109/CVPR.2001.990517.

[18] Wu, J., Jenerette, G.D., Buyantuyev, A. and Redman, C.L. 2011. Quantifying spatiotemporal patterns of urbanization: The case of the two fastest growing metropolitan regions in the United States. *Ecological Complexity*. 8, 1 (Mar. 2011), 1–8. DOI:https://doi.org/10.1016/J.ECOCOM.2010.03.002