WinSet: The First Multi-Modal Window Dataset for Heterogeneous Window States

Tzu-Yi Fan, Tun-Chi Tsai, Cheng-Hsin Hsu
Department of Computer Science
National Tsing Hua University
Hsin-Chu, Taiwan

Fanqi Liu, and Nalini Venkatasubramanian
Department of Computer Science
University of California
Irvine, USA

Abstract

Windows play an important role in modern buildings. Getting to know the window states, e.g., open vs. close, is an enabler of many smart city applications, such as energy conservation and emergency response. In this work, we collect the very first multi-modal (RGB, thermal, depth, LiDAR, and ultrasound) window dataset named *WinSet* at various distances and angles. Multiple window types and heterogeneous window states are considered, such as openness (open vs. close), human behind (with vs. without), and lighting (on vs. off). Although our WinSet dataset has many usage scenarios, we concretize two sample ones: (i) analysis of state distinguishability using different sensor modalities and (ii) algorithms to detect open windows. We believe sharing WinSet and its collection procedure with the engineering and research communities will stimulate many creative smart city applications.

CCS Concepts: • **Information systems → Multimedia information systems**.

Keywords: dataset, sensors, semantic segmentation, window localization, window state detection, machine learning

ACM Reference Format:
Tzu-Yi Fan, Tun-Chi Tsai, Cheng-Hsin Hsu and Fanqi Liu, and Nalini Venkatasubramanian. 2021. WinSet: The First Multi-Modal Window Dataset for Heterogeneous Window States. In *Proceedings of The 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys), November 17–18, 2021, Coimbra, Portugal.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3486611.3486662

BuildSys '21, November 17–18, 2021, Coimbra, Portugal

ACM ISBN 978-1-4503-9114-6/21/11...$15.00
https://doi.org/10.1145/3486611.3486662

1 Introduction

Urban and suburban settings consist of building structures that vary in size, height, age, and use. Today, Building Management Systems (BMS), integrated with Heating, Ventilation, Air Conditioning (HVAC) and surveillance systems, are installed in large public or commercial spaces for better occupant comfort, security, and safety. In this paper, we collect a dataset focusing on a critical structural aspect of all urban buildings: *windows*. Windows are keys to several building functions, and their states influence not only aspects of ventilation and lighting [17] but also energy consumption, public safety, and community management. For example, if windows are left open, air conditioners may consume more electricity resulting in excessive greenhouse gas emissions [18], while thieves may sneak into the building causing security concerns. At a fire scene, the extra oxygen and wind from an open window may fuel the fire or change its spreading pattern, and cause complications to firefighters. In addition, humans may appear at the windows when trying to leave or escape from the buildings. For example, the police department may want to monitor all windows to locate active shooters or depressed citizens. Moreover, during epidemic time, governments increasingly promote the benefits of opening windows of crowded rooms to reduce the transmission of tuberculosis [6].

Therefore, getting to know the *locations* and *states* of windows is crucial to many smart city applications. Examples of the window states include: *openness* (open vs. close), *human* behind (with vs. without), and *lighting* (on vs. off). One way to determine the window states is to install dedicated sensors, such as magnet switches for window openness. However, it costs a lot for the deployment and maintenance, so dedicated sensors are not installed by every building (especially in developing countries). Furthermore, most dedicated sensors can only detect one window state, and we may not have access to the sensor data when emergency occurs. To cope with such limitations, *image* and other *rich-media* sensors can be adopted. These sensors can be set up at the fixed positions for persistent monitoring, or carried by humans or robots/drones for dynamic monitoring.

Detecting window states using rich-media sensors is inherently challenging. For example, it is hard to detect window openness using RGB cameras because the light reflection and transmission in glass are complicated processes affected by too many factors. Therefore, from different angles, RGB cameras may see through the window glass or observe reflected views. Fortunately, there exist many rich-media sensors for diverse *modalities* other than visible lights captured by RGB cameras. Nonetheless, most public datasets [4] are collected by RGB cameras only. Hence, it is no easy task to develop multi-modal classification algorithms for window states.

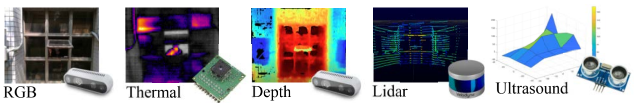


Figure 1. Considered sensor modalities.

In this paper, we collect the first multi-modal window state dataset named *WinSet*[1] for novel smart city applications. We employ the following sensor modalities: RGB, thermal, depth, LiDAR, and ultrasound, as illustrated in Fig. 1. WinSet dataset consists of two parts: datasets A and B. Dataset A collects the multi-modal window images with distinct states from different angles and at diverse distances. Possible usages of this dataset include finding the best sensor combination for detecting each window state, and designing concrete state classifiers. A classifier for human behind, for example, could be a part of BMS [3], applying data to help victim rescue in firefighting, or used by the police department to locate active shooters or prevent suicide cases. Dataset B collects the multi-modal images of various types of windows with different openness states at a given angle (0°) and distances (1, 2, and 3 m). One possible usage of this dataset is designing state classifiers that work across different window types. *To the best of our knowledge, WinSet is the very first multi-modal window dataset, while the collection procedure of similar datasets has never been proposed in the literature.*

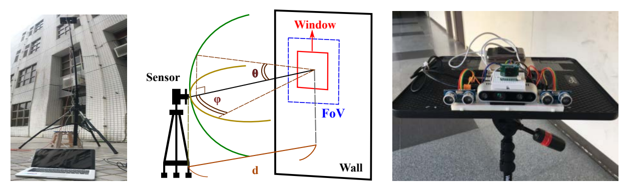
2 Related Work

To our best knowledge, there is only one dataset focusing on the window openness state. Particularly, Safavi et al. [19] collected a sound dataset with four window states: open, close, open-to-close, and close-to-open. Their dataset concentrated on sound, and thus is quite different from WinSet. While not exercising diverse window states, RGB datasets on windows have been collected for window localization [15]. Examples of such RGB datasets include Gadde et al. [7], and Daftry et al. [5]. Besides RGB, datasets from other modalities were also collected. For example, Malihi et al. [16] employed LiDAR to create building structure datasets, while Sirmacek et al. [20] adopted a thermal camera to capture a building opening dataset. The above-mentioned window-related datasets only considered a single modality, and thus is different from WinSet. There were also attempts on combining two sensor modalities, e.g., Jarzabek et al. [12] and Lin et al. [14] captured RGB and thermal images of buildings for detecting air leaks to save energy.

[1]Please contact the authors for the access to the dataset.

Table 1. Specifications of Adopted Sensors

A	B	Modality	Make/Model	Technology	Sampling Rate	Field of View	Raw Format
✓	✓	RGB	Intel Realsense D435	Visible Light	680×480 @15 fps	69°×42°	`.bag`
✓		Thermal	FLIR Lepton 2.0	Long Wavelength Infrared	80×60 @8.6 fps	50°×40°	`.bin`, `.png`
	✓	Thermal	FLIR Lepton 3.5	Long Wavelength Infrared	160×120 @8.7 fps	57°×45°	`.bin`, `.png`
✓	✓	Depth	Intel Realsense D435	Active Stereo Infrared	680×480 @15 fps	86°×57°	`.bag`
✓		LiDAR	Velodyne Puck VLP-16	Laser	1875×16 @10 fps	360°×30°	`.pcap`
	✓	Ultrasound	OSEPP HC-SR04	Sonar	1×1 @10 fps	60°×60°	`.txt`



(a) (b) (c)

Figure 2. Data collection setup: (a) 7-meter tripod, (b) tripod setup for dataset A, and (c) sensor setup for dataset B.

3 Collection Procedure

3.1 Dataset A

Hardware. We choose four off-the-shelf sensors to collect dataset A, as reported in Table 1. The RGB camera captures visible lights into images. The thermal camera uses a microbolometer to detect the infrared (emitted by objects), translates it into relative temperature, and creates thermal images. The depth camera collects two parallel images, and then calculates the depth of each pixel. Last, LiDAR emits laser beams, assesses the duration for the beams to bounce back, calculates the distances to objects, and generates a 3D map for the surrounding environment. Because we plan to collect dataset A from systematically selected angles, we adopt a 7-meter tripod with a pan-tilt head, as illustrated in Fig. 2(a). We attach all sensors on a platform, which is then mounted on the tripod head.

Software. We use software tools provided by the sensor manufacturers to capture image frames. For example, we install Intel Realsense Viewer from their official GitHub [11] on our laptop to record RGB and depth image frames. The tool compresses each pair of RGB and depth image frames into a `.bag` file. For the thermal sensor, we set it up on a

Raspberry Pi 3, and compile the code of Lepton module from Groupgets [9] to get thermal image frames. We also save the original thermal data using a Python script into `.bin` files. For LiDAR, we utilize VeloView [13] to capture the detected 3D points into `.pcap` files, frame by frame.

Steps. We collect data from four exterior windows on our campus. We vary the following parameters: (i) distance $d \in \{3, 6, 12\}$ meter, (ii) polar angle $\theta \in \{0°, 30°\}$, and (iii) azimuthal angle $\varphi \in \{0°, 30°, 60°\}$, as shown in Fig. 2(b). We consider three representative window states: (i) openness, (ii) human, and (iii) lighting. Each measurement lasts for 10 seconds.

3.2 Dataset B

Hardware. We also choose four sensors for dataset B, as reported in Table 1. Compared to dataset A, we use a thermal camera with a higher resolution of 160 x 120. We employ two ultrasound sensors. Both exterior and interior windows are considered in dataset B. We set up all sensors on a platform, which is revealed in Fig. 2(c). We set two ultrasound sensors at the two sides of the Intel Realsense, inspired by Bai et al. [1]. This is to detect the two casements that are common among several window types. All sensors are connected to a Raspberry Pi 4 for better mobility.

Software. For RGB and depth sensors, we use Intel Realsense rs-capture instead of Intel Realsense Viewer to conserve energy of Raspberry Pi 4. Different from other sensors, the ultrasound sensors only get readings rather than images. We program the ultrasound sensors to detect ten times every second, and save the values into a `.txt` file.

Steps. We consider diverse window types, such as sliding, awing, barred, and screened windows. We point the sensors to the center of each window (both θ and φ are 0°), and collect the dataset at three different distances $d \in \{1, 2, 3\}$ m. Each measurement lasts for 10 seconds. We focus on the openness state, and only keep 20 image frames for each measurement in the dataset.

3.3 Semantic Labeling

We annotate each image pixel using Labelme [21] with one or multiple labels, including glass, window frame, wall, floor, ceiling, human, open window, and background. Because RGB images have the highest resolution, we decide to label RGB images only. Moreover, for each RGB video, we only label 20 equal-distanced sample image frames to leverage temporal redundancy.

4 Sample Usage Scenarios

4.1 Distinguishability of Different Sensors

To understand which sensors have better distinguishability among different window states, we analyze the image frames from dataset A. In the following discussion, we take window openness as an example. Other window states can be analyzed with the same approach. For any two images, we employ two metrics to quantify their distinguishability: (i) *Histogram Correlation (HC)* [2] and (ii) *Number of Gaussian till Homogeneity (NG)* [8]. Lower HC, as well as higher NG, values mean higher distinguishability.

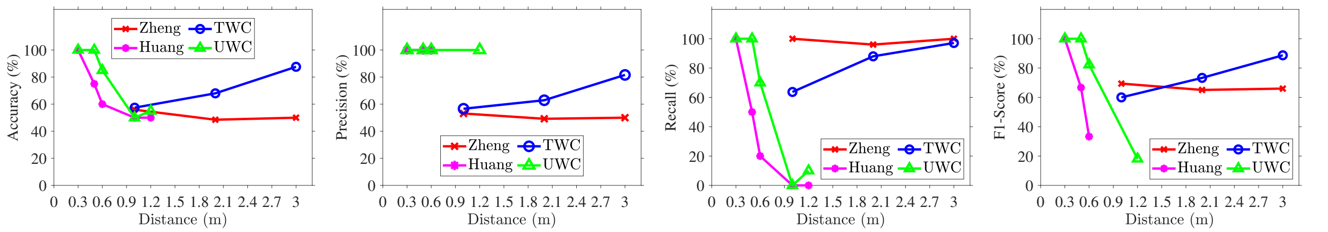
To compute statistically meaningful results, for each window, distance, and azimuthal angle, we randomly select 100 pairs of images with *different* window openness states from dataset A, and 100 pairs of images with the *same* window openness states (both open or close) to compute their distinguishability individually. A good sensor modality should show high distinguishability with different window states, and low distinguishability with the same window states. We make the following key observations on our experiment results[2]: (i) depth images don't work at all, (ii) both thermal and RGB images have good open window distinguishability, (iii) RGB images have the best performance at all distances, and (iv) sensors work the best at 0°and the shortest distance; the largest distinguishability appears at 3 m and 0°. Hence, we conclude that the most promising sensor modalities for distinguishing window openness are: RGB and thermal sensors. *This usage scenario demonstrates that WinSet can be analyzed for novel insights of heterogeneous smart city applications using different sensor modalities.*

4.2 Open Window Detection

We design two open window detection algorithms using dataset B. We use thermal images to develop Thermal Window Classification (TWC) algorithm based on the following intuition. We expect to observe the consistent temperature from the whole window when a window is close, and diverse temperature distribution when a window is open. TWC generates a histogram of all pixels. If there exists only one local minimum in the histogram, we declare the window is open, because there are two clusters of thermal readings. We also develop an Ultrasound Window Classification (UWC) algorithm, which compares the distance returned by the ultrasound sensor against the actual distance to a window (e.g., derived from the GPS coordinates or building blueprint). If the ultrasound sensor data is larger than the actual distance by a cushion parameter (20 cm if not otherwise specified), we declare the window is open.

We have also implemented two existing algorithms as our baseline algorithms: Zheng et al. [22] and Huang et al. [10]. Fig. 3 compares the performance of the algorithms at different distances. We make the following key observations. First, the ultrasound-based algorithms (Huang and UWC) perform worse at longer distances. However, our proposed UWC constantly outperforms Huang once the distance goes above 0.5 m. In terms of recall, Huang and UWC drop to close to 0 at 1 m. Second, the RGB/thermal-based algorithms (Zheng and TWC) perform better at longer distances. Furthermore, our proposed TWC consistently outperforms

[2]Figures are omitted due to space limit.



(a) (b) (c) (d)

Figure 3. The performance of our two open window detection algorithms at different distances: (a) accuracy, (b) precision, (c) recall, and (d) F1-score.

Zheng in both accuracy and F1-score, and the performance gap increases as the distance is increased. Based on this figure, we recommend the UWC algorithm for fine-grained detection (≤ 40 cm), and the TWC algorithm for coarse-grained detection (> 40 cm). *This usage scenario demonstrates that WinSet can be used to develop and evaluate window state classifiers for smart city applications.*

5 Conclusion

We have collected a multi-modal window dataset, called WinSet. WinSet consists of two datasets: (i) dataset A for window images with distinct states from different angles at various distances, and (ii) dataset B for window images of different window types with different states at diverse distances. To the best of our knowledge, there exists no prior dataset collected with so many modalities, states, distances, and angles. We carefully organize the data files and provide alternative file formats to ease the burden on the engineers/researchers who want to use our dataset. WinSet can be used in many usage scenarios, including but not limited to what we showed: sensor data analysis and window state classifiers. In addition to our dataset, the data collection procedure can be adopted by the research community. We highly welcome contributions from all over the world since windows are quite different among continents.

Acknowledgement

The authors thank Yan-Mei Tang and Yi-Ting Wang for their help on sensor data collection; T. Fan was supported by a Novatak Fellowship; the research is partially funded by the Ministry of Science and Technology (Taiwan) (#110-2221-E-007-102), Defense Advanced Research Projects Agency (#FA8750-16-2-0021), and UC Office of the President (#LFR-20-653572).

References

[1] J. Bai, S. Lian, Z. Liu, K. Wang, and D. Liu. 2017. Smart Guiding Glasses for Visually Impaired People in Indoor Environment. *IEEE Transactions on Consumer Electronics* 63, 3 (August 2017), 258–266.

[2] E. Bazan, P. Dokladal, and E. Dokladalova. 2019. Quantitative Analysis of Similarity Measures of Distributions. In *Proc. of British Machine Vision Conference (BMVC'19)*. Cardiff, United Kingdom.

[3] K. Benson, G. Wang, N. Venkatasubramanian, and Y. Kim. 2018. Ride: A Resilient IoT Data Exchange Middleware Leveraging SDN and Edge Cloud Resources. In *Proc. of IEEE/ACM International Conference on Internet-of-Things Design and Implementation (IoTDI'18)*. Orlando, FL, 72–83.

[4] D. Ceylan, N. Mitra, Y. Zheng, and M. Pauly. 2014. Coupled structure-from-motion and 3D symmetry detection for urban facades. *ACM Transactions on Graphics (TOG'14)* 33, 1 (January 2014), 1–15.

[5] S. Daftry, C. Hoppe, and H. Bischof. 2015. Building with drones: Accurate 3D facade reconstruction using MAVs. In *Proc. of IEEE International Conference on Robotics and Automation (ICRA'15)*. Seattle, WA, 3487–3494.

[6] R. Escombe, E. Ticona, V. Chavez-Perez, M. Espinoza, and D. Moore. 2019. Improving natural ventilation in hospital waiting and consulting rooms to reduce nosocomial tuberculosis transmission risk in a low resource setting. *Springer BMC infectious diseases* 19, 1 (January 2019), 1–7.

[7] R. Gadde, R. Marlet, and N. Paragios. 2016. Learning grammars for architecture-specific facade parsing. *Springer International Journal of Computer Vision* 117, 3 (March 2016), 290–316.

[8] R. Gonzalez and R. Woods. 2018. Filtering in the Frequency Domain. In *Digital Image Processing*. Pearson, Chapter 4, 270–271.

[9] Groupgets. 2019. LeptonModule. https://github.com/groupgets/LeptonModule.

[10] Z. Huang, K. Wang, K. Yang, R. Cheng, and J. Bai. 2018. Glass Detection and Recognition Based on the Fusion of Ultrasonic Sensor and RGB-D Sensor for the Visually Impaired. In *Proc. of SPIE Target and Background Signatures IV*. Berlin, Germany, 118–125.

[11] Intel. 2021. Intel Realsense SDK. https://github.com/IntelRealSense/librealsense.

[12] M. Jarzabek, D. Lin, and H. Maas. 2020. Supervised Detection of Facade Openings in 3D Point Clouds with Thermal Attributes. *MDPI Remote Sensing* 12, 3 (February 2020), 543.

[13] Kitware and Velodyne. 2014. VeloView. https://www.paraview.org/VeloView.

[14] D. Lin, Z. Dong, X. Zhang, and H. Maas. 2019. Unsupervised Window Extraction from Photogrammetric Point Clouds with Thermal attributes. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences* 4 (June 2019), 10–14.

[15] H. Liu, Y. Xu, J. Zhang, J. Zhu, Y. Li, and S. C. H. Hoi. 2020. DeepFacade: A Deep Learning Approach to Facade Parsing With Symmetric Loss. *IEEE Transactions on Multimedia* 22, 12 (December 2020), 3153–3165.

[16] S. Malihi, M. Valadan Zoej, M. Hahn, and M. Mokhtarzade. 2018. Window Detection from UAS-Derived Photogrammetric Point Cloud Employing Density-Based Filtering and Perceptual Organization. *MDPI Remote Sensing* 10, 8 (August 2018), 1320.

[17] G. Menzies and J. Wherrett. 2005. Windows in the workplace: examining issues of environmental sustainability and occupant comfort in the selection of multi-glazed windows. *Elsevier Energy and buildings* 37, 6 (June 2005), 623–630.

[18] One Source Heating and Cooling. 2016. How Does Opening Your Windows Impact Your HVAC System? https://www.onesourceair.com/blog/opening-windows-impact-hvac-system.

[19] S. Safavi, T. Iqbal, W. Wang, P. Coleman, and M. Plumbley. 2020. Open-Window: A Sound Event Data Set For Window State Detection And Recognition. In *Proc. of International Workshop on Detection and Classification of Acoustic Scenes and Events (DCASE'20)*. Tokyo, Japan.

[20] B. Sirmacek, L. Hoegner, and U. Stilla. 2011. Detection of Windows and Doors from Thermal Images by Grouping Geometrical Features. In *Proc. of IEEE Joint Urban Remote Sensing Event (JURSE'11)*. Munich, Germany, 133–136.

[21] K. Wada. 2016. labelme: Image Polygonal Annotation with Python. https://github.com/wkentaro/labelme.

[22] H. Zheng, F. Li, H. Cai, and K. Zhang. 2019. Non-Intrusive Measurement Method for the Window Opening Behavior. *Elsevier Energy and Buildings* 197 (August 2019), 171–176.