ストリートビュー画像を用いた都市環境の 3D 再構築とセグメンテーション分析法

レン　モニーカンディ*・顧　宗超**・曹 俊**

Reconstructing 3d Segmentation of the Urban Environment using Street View Images

Monykandy Leng*, Zongchao Gu**, Jun Cao**

Abstract: The recreation of the environment can be a complex process involving a lot of resources and time to produce. Even if the city has been produced digitally, the maintaining and updating process would require continuous investments. The study proposes a method to produce the 3-dimensional spatial relations using only available street view images as the source of the input. First, the images are pre-process before producing the 3d environments, The resulted point cloud is later combined with segmentation of the images producing a the segmented 3d environments. This results not only the recreation of the environments but its segmentation can also be used in other urban study purposes, such as seismic risk assessment, simulations, or data recreation for mid to small scale study of the urban area. The researches introduce a tool to easily obtain the spatial relations of the city using minimal resources and can be used for other urban study purposes.

Keywords: 3D セグメンテーション（3d segmentation），3D 再構築（3d reconstruction），パノラマ画像（panoramic images），ストリートビュー画像（street view images）

1. Introduction

Street View images has become one of the important tools used in urban analysis. However, the studies are grounded into image analysis from the camera mounted on the vehicle, implying that the analysis usually consider the 3-dimension space as if it was flattened into a surface, viewed in 360 degrees but from one point of view that is located at the street. With the advancement of the 3d recreation tools, the 360 images captured from a vehicle with around 10 meters interval can be used to recreate the 3d environment and its 3d segmentation (Figure 1) providing 2 outputs that can be used for different purpose of analysis.

The introduction of the tool to create virtual environment would enable the study of the urban area to be in an environment that is, supposedly more realistic than flattened images. The built 3d environment would allow for various analysis such as, creating simulations, generating images from other point of views or the estimation of the shaded area. Furthermore, the segmentation of the 3d environment would allow for assessment of the seismic risk of the space, the study of green volumes in relation to perceived spaces (attractiveness, safety etc.,) as the city is experienced, or the wind flow and air quality analysis in a mid to small scale.

While it is true that the data of the city can be built manually or has existed in some place already, the large scale 3d reconstruction often overlook these small details and changes which could affect the analysis if the case study is done in a smaller scale. In addition, the data of the built objects and its identification requires the understanding of, if not accurately reconstructed environment, the 3-dimensional spatial relation of the environment. In addition, some of these data are not static, for instance, the growth of the tree in the relation to its age, the age of the building and the building extension such as arcade, or air conditioning unit could be changed at any given moment. Therefore, the reconstruction of 3d environment in a traditional way would not only require a lot of time and resource investment to create, but maintaining its accuracy over time would be resource consuming as well.

The paper proposes a method to reconstruct 3d environment from a series of 360 images, using minimal resources to obtain the similar result to the reality, and creating the segmentation of the 3d environment, hence, creating a data that is easier and has more data (3 dimensionality) to interpret.

2. Methods

For each segment of the road, the 360 images are collected every 10 meters to prepare for the 3d reconstruction. After obtaining the 360 images similar to SVIs, the process are followed as shown in Figure 2. The images are enhanced, if

*　学生会員　東南大学建築学院（Southeast University）
〒210096　中國南京市四牌楼 2 号　Tel：13921439438，E-mail：monykandyleng@gmail.com
**　非会員　　東南大学建築学院（Southeast University）

needed, to ensure the image has good resolution. Next, the 360 images are split into 4 perspective images with the field of view of 120 and the camera intrinsic is calculated and written in ".yaml" format for reconstruction as it is required as an input for the reconstruction. The calculation can be done alongside the image transformation as its parameters are controlled when converting 360 images into perspective images. Then the reconstruction is run using Monocular Surface Priors for Robust Structure-from-Motion (MP-SfM) (Pataki et al., 2025) to get the 3d environment. Later, the data of the 3d point clouds is extracted from the output of the 3d reconstruction and combined with the segmentation data from the images to produce the 3d environment with segmentation.

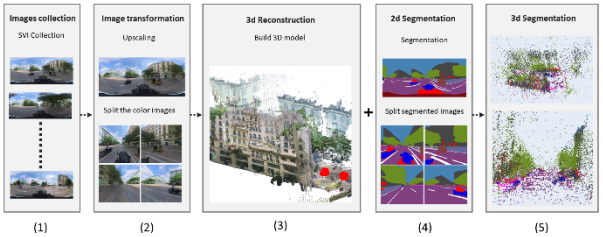


Figure 2. Research workflow: 1) images collection; 2) image transformation; 3) 3d reconstruction; 4) 2d segmentation; 5) 3d segmentation

2.1 Data collection

The 360 images are collected in an island called Koh Pich (Diamond Island) located in Phnom Penh, Cambodia, using Insta360 X3 with the resolution of 5760 x 2880 pixels with 96 dpi. The distance between each 360 images is approximately 10 meters apart. The studied length is a road segment of 70 meters in total in a compact area with buildings along the road (Figure 3.). The site is selected for its variation of shapes, curvature and distance of the buildings from the street.

However, the current Street View Images are not up to date and the quality of the image may not need the required resolution for a successful reconstruction. Therefore, the data collection process has to be done manually. To get the image intervals as close to the SVIs data as possible (10-15 meters from one to another, taken on the road with cameras on top of the car), the set-up is as follows:

1) The 360 camera is mounted on a motorbike at the height of approximately 2 meters from the ground.
2) The vehicle is driven at almost constant speed 30km/h
3) The video is set to capture at 30 frames per second
4) The time where the vehicle accelerates and decelerate is cropped out leaving only 2-3 spots where the speed of the vehicle is slowed down to the minimum of 26 km/h
5) After calculation, the image is extracted every 36 frames which should be 10 meters apart or less if the vehicle slows down.

The collected images are calibrated to ensure the north point stays on the centerline all across images.

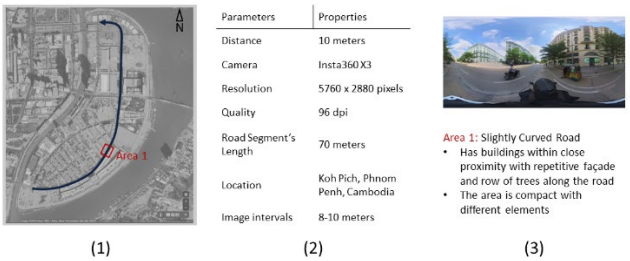


Figure 3. Data collection: 1) reconstruction site; 2) camera parameters and images information; 3) characteristic of the selected area

2.2 Image transformation

Before processing the images for 3d reconstruction, a few factors need to be considered in order to ensure there is no problems during the reconstruction.

Firstly, the quality of the images has to be ensured to be high, as the low-quality images could result in fragmented point clouds. In this case, the 360 images used is 5760 x 2880 pixels. If the requirements is not met, Clarity Refiners UI is used from Pinokio to upscale the images dimensions and enhance its quality.

Second, the 360 images are transformed into 4 perspective images with the dimension of 1000 x 1000 pixels with 96 dpi and the Field of View (FOV) 120 to ensure that there are enough overlapping percentage from one image to another. Alongside the transformation, the camera's intrinsic is calculated and saved in a .yaml format. This is because, the reconstruction requires this file to operate. Since the process of transforming 360 images to perspective image is essentially a reprojection of sphere being projected to flat surfaces, and the reprojection parameters are controlled using the tools from ZenSVI (Ito et al., 2024), the data for camera's intrinsic can be calculated and written for the 3d reconstruction.

At the same time, the 360 images are segmented using Mapillary and the tools from ZenSVI (Ito et al., 2024) as well. Then the segmented images are transformed into perspective using the same tools and parameters of the normal images for later use.

2.3 3d Reconstruction

The reconstruction of 3d environment from images has been introduced for quite some time with a variation of tools for different output including photogrammetry, Gussian Splat, NeRFs, however, these methods require huge overlaps, meaning the intervals between images has to be small which would require a big dataset and a lot of processing time to create the 3d environment (Fang et al.,2025; Murtiyoso et al.,2023) (Wu, et al., 2025).

A new method called Monocular Surface Priors for Robust Structure-from-Motion, (Pataki et al., 2025) has enhanced the 3d reconstruction pipeline to a point where the overlapping percentage between images required for the process is so low that we can use images 10 meters apart like the data from SVIs. Hence, the tool was used to re-create the environment and explore its potential to be used in urban analysis. The recreation was run on the default configuration sp-lg_m3dv2 with verbose 2 to get the fast computation time and the visualization output as well. The result is an .html file with point clouds of the 3d environment and the estimated camera's position as shown in Figure 5.

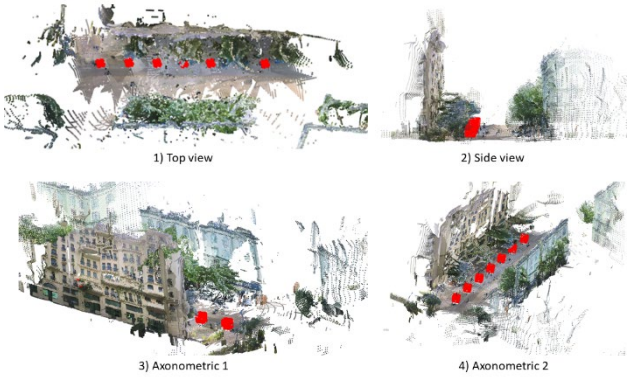


Figure 5. Reconstructed point clouds: 1) Top view; 2) Side view; 3) Axonometric 1; 4) Axonometric 2

2.4 3d Segmentation

Upon finishing the 3d reconstruction the information is extracted and combined with the segmented image. The color of the point cloud can be re-applied using segmentation color at that 3d point position. This will create the 3d point clouds using the segmentation color instead of normal color (Figure 6.).

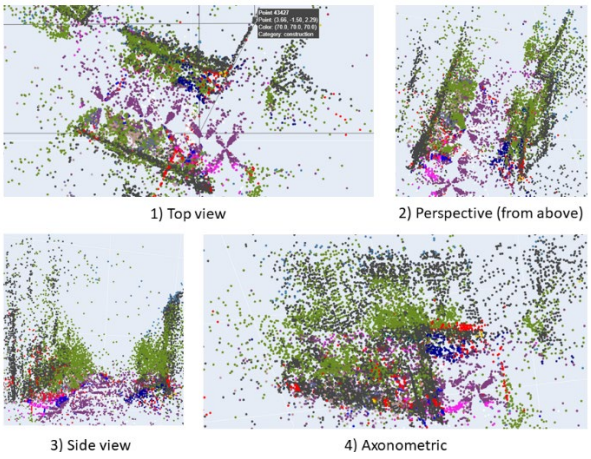


Figure 6. 3d Segmented point clouds: 1) Top view; 2) Perspective (from above); 3) Side view; 4) Axonometric

The segmented 3d points cloud essentially provides two different types of data in a single format, that is segmentation and spatial relation of the scene. Therefore, its usage can be applied in various field. For instance, in the seismic risk assessment (Section 3.1), the segmented point clouds can be used as building and depth information or a land cover map. It can also be used to generate new data including new image generation, spatial studies of the urban fabric, or mesh generation.

3. Discussion

3.1 Potential application in seismic risk assessment

The seismic activity is one of the natural occurrences that cannot be prevented, and the preparedness towards the resilience of the city is crucial to minimize its impact. The methods introduced allows for the reconstruction and identification of the urban scene providing the instruments to study the potential risk zones and estimate the safe circulation path for evacuation (Gu et al.,2024). The workflow can be used in the following proposed diagram (Figure 7.)

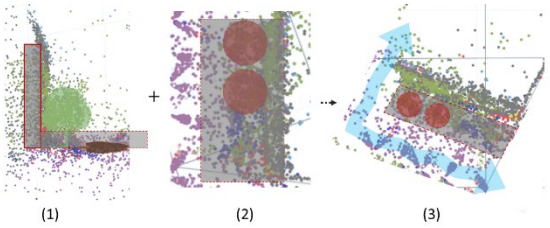


Figure 7. Safe zone identification: 1) risk region identification; 2) landcover map; 3) safe zone assessment

3.2 Potential application in solar exposure simulations

The solar exposure estimation in the urban scene, traditionally, relies on the image processing from one point of view which normally is the one taken from the vehicle side. With the introduction of point clouds 3d reconstruction, the solar exposure can be studied using panorama images (Zhu & Gu, 2022) in a pedestrian path or any other points that is not from the image input. The new generated image could be used at that new point of view to produce a sky view factor and estimate the solar exposure from the specified point directly (Figure 8.). With the analysis being completed, the suitable strategies and implementations can be explored by related urban actors to provide a comprehensive solution specific to that area.

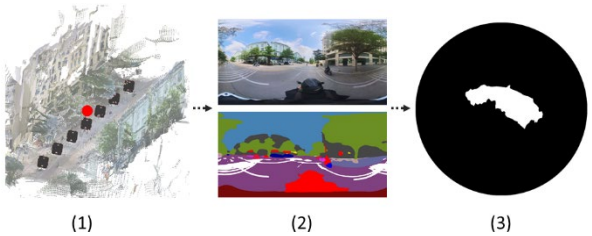


Figure 8. Solar exposure analysis: 1) a new camera position (red dot) can be picked; 2) new 360 image and segmentation can be produced; 3) sky view factor is generated

3.3 Potential application in thermal comfort analysis

The exploration of the usage potential of the segmented 3d environment can be expanded as well. The 3d point clouds can be used as an instrument to obtain the 3d mesh. The resulted mesh can be combined with on-site data to run the simulation to estimate thermal comfort (Chen et al., 2021) (Figure 9.).

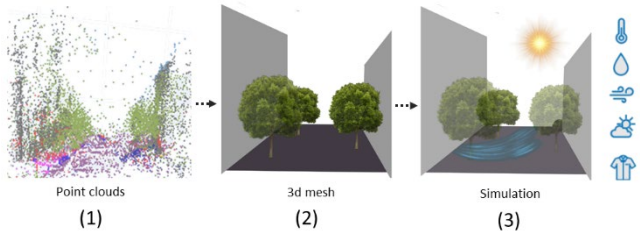


Figure 9. Thermal comfort analysis: 1) point clouds output data; 2) construct the 3d model; 3) combine with other on-site data for simulations to estimate the thermal comfort

The usage of the 3d environments and its segmentation can be explored and applied in other urban studies as well. The proposed method allows for 3d reconstruction using only a small number of a common data, street view images, and estimates the spatial relation to be used in different scenarios.

3.4. Limitations

The proposed method has a number of potential error as it was requiring only a small amount of data. However, the rooms for error can be identified.

1) The reconstruction that has buildings located considerably far from the road produce too few points cloud which, sometimes, cannot be interpreted.
2) The area that has sharp curves might result in the case in which algorithm cannot detect the movement of the camera and simply put multiple cameras in the same position producing the wrong points cloud.
3) The segmentation is not using all of the points cloud in the 3d reconstruction.
4) There are too many points cloud to be used in urban scale, therefore, the optimization is needed.
5) The points cloud demonstrates only the spatial relation of the space, so its scale with the real-world has to be confirmed

From these limitations, it is suggested to collect data with close proximity especially in area with sharp turns and not much distinctive features visible nearby, as it can be challenging to reconstruct the 3d environment. Its potential application, format conversion, and real scale can be further studied.

However, it is worth noting that the tested reconstruction uses only the default configuration for fast and light computation process rather than the heavy one. Therefore, the result may differ according to the configuration setting as well.

4. Conclusion

The study proposes a method to obtain the data of the spatial relation of the environment and segment the point clouds in the 3-dimension using only panoramic images 10 meters apart. The constructed point clouds and segmentation can be used to understand the urban scenes and experiences allowing for potential application in urban studies such as circulation space assessment, simulations, thermal comfort studies etc.

References

Chen, G., Rong, L., & Zhang, G. (2021). Unsteady-state CFD simulations on the impacts of urban geometry on outdoor thermal comfort within idealized building arrays. *Sustainable Cities and Society*, *74*, 103187. https://doi.org/10.1016/j.scs.2021.103187

Fang, X., Zhang, Y., Tan, H., Liu, C., & Yang, X. (2025). Performance Evaluation and Optimization of 3D Gaussian Splatting in Indoor Scene Generation and Rendering. *ISPRS International Journal of Geo-Information*, *14*(1), 21.

Gu, Z., Sun, X., & Zhu, H. (2024, October). *A simple method of measuring the seismic risk of circulation space in urban residential streets based on street-view data*. Paper presented at the Annual Meeting of the Geographic Information Systems Conference.

Ito, K., Zhu, Y., Abdelrahman, M., Liang, X., Fan, Z., Hou, Y., Zhao, T., Ma, R., Fujiwara, K., Ouyang, J., Quintana, M., & Biljecki, F. (2024). ZenSVI: An Open-Source Software for the Integrated Acquisition, Processing and Analysis of Street View Imagery Towards Scalable Urban Science. *Comput. Environ. Urban Syst., 119*, 102283.

Murtiyoso, A., & Grussenmeyer, P. (2023). INITIAL ASSESSMENT ON THE USE OF STATE-OF-THE-ART NERF NEURAL NETWORK 3D RECONSTRUCTION FOR HERITAGE DOCUMENTATION. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLVIII*(M-2-2023), 1113–1118.

Pataki, Z., Sarlin, P. E., Schönberger, J. L., & Pollefeys, M. (2025). MP-SfM: Monocular Surface Priors for Robust Structure-from-Motion. In *Proceedings of the Computer Vision and Pattern Recognition Conference* (pp. 21891-21901).

Wu, S., Feng, L., Zhang, X., Yin, C., Quan, L., & Tian, B. (2025). Optimizing overlap percentage for enhanced accuracy and efficiency in oblique photogrammetry building 3D modeling. *Construction and Building Materials*, *489*, 142382. https://doi.org/10.1016/j.conbuildmat.2025.142382

Zhu, H., & Gu, Z. (2022). A method of estimating the spatiotemporal distribution of reflected sunlight from glass curtain walls in high-rise business districts using street-view panoramas. *Sustainable Cities and Society*, *79*, 103671. https://doi.org/10.1016/j.scs.2022.103671