

3D Modeling of Building Façade from a Single Distorted Street-View Panorama

Qing Zhu, Mier Zhang, Han Hu and Feng Wang

Abstract—Façade features are important towards large-scale LoD-3 reconstruction in urban environment and street-view panoramas are arguably the best option for detailed 3D façade modeling. However, despite the plethora of street-view panoramas in the community, it is surprising that few researches have been devoted to explore the metric capabilities of the panoramas. This is by part due to the complexities in system integration and by part the projection (e.g. distortion at the building tops) and deformation (e.g. the straight structures being bended) problems of the panoramas. Aiming to solve the above issue, this paper introduces a flexible and practical solution using only a single panorama. The key is to efficiently rectify the panoramas using image-space line constrained deformation inspired by the as-rigid-as-possible deformation of the surface meshes. In addition the image is also re-projected using gnomonic projection on a properly selected tangent plane. Experimental evaluations have revealed the effectiveness of the image-space rectification; after proper scaling, the semantic-aware 3D façade models have achieved decimeter-level accuracies with respect to the reference surface meshes.

Index Terms—3D Modeling, Panorama, Façade, Distortion Rectification

I. INTRODUCTION

The 3D modeling of the building façades in urban environment is part of the LoD-3 (Level-of-Details) reconstruction [1] and an important prerequisite for various applications ranging from the community of urban design and disaster management [2] to the industry of entertainment such as virtual reality and the creation of movies [3]. Unlike the traditional nadir images designed for building roofs [4], aerial oblique images, mobile laser scanning and street-view panoramas are the most common datasets that have the capability to observe the building façades. However, aerial oblique images suffer severe occlusion problem, limited spatial resolution and perspective deformation [5] and mobile laser scanning is confined to the inhomogeneity of the point samples and the incompleteness of the data acquisitions, which is more suitable for features of the roads [6] rather than façades. Therefore, street-view

Manuscript received: August 31, 2019. This work was supported in part by the National Natural Science Foundation of China (Projects No.: 41631174, 6102392, 41871291) and in part by the National Key Research and Development Program of China (Project No.: 2018YFB0505404). (Corresponding author: Han Hu)

Qing Zhu, Mier Zhang, and Feng Wang are with the Faculty of Geoscience and Environmental Engineering, Southwest Jiaotong University, Chengdu, China. (e-mail: zhuq66@263.net; mierzhang@my.swjtu.edu.cn; wangfeng19920903@my.swjtu.edu.cn)

Han Hu is with State Key Laboratory of Rail Transit Engineering Informationization, China Railway First Survey and Design Institute Co. Ltd., Xi'an, China and also with the Faculty of Geoscience and Environmental Engineering, Southwest Jiaotong University, Chengdu, China. (email: han.hu@swjtu.edu.cn)

panoramas, because of the 360° field of view and the widely existing community availability [7], are arguably the best option for detailed façades modeling.

With regard to 3D reconstruction, Micusik *et al.*[8] produced street-view textured mesh models using panoramas, but the output was represented as irregular triangles and not suitable for certain applications that required lightweight low-polygon models; similar approach is also adopted to community photo collections [9]. In addition, many works required more auxiliary data that were not flexible or even not applicable using the community street-view panoramas. For instance, some approaches began with the original perspective images that were not distributed by most street-view vendors [10], [11]; [12] even required the use of the depth sensor (also known as RGB-D camera) to be attached with the panorama and because of the limited depth range (3 to 5 m in general), it was strictly limited to indoor applications; the 3D depth sensor could be substituted with laser scanners [13], an LoD-2 model [14] or radar [15], but this requirement was still beyond the use of a single panorama and reduced the flexibility of the system design.

Even though the availability of panoramas are ready and most 3D modeling software support the import of reference images [16], we have yet seen many solutions on this topic. This is partly because the street-view panoramas have some defects, as shown in Figure 1: (1) *Deformation of the images*. Street-view panoramas are generally mosaicked from multiple cameras. In theory, if the relative orientation between different cameras and the intrinsic parameters of the cameras were calibrated perfectly and the horizontal direction of the panoramas leveled strictly, the vertical lines on the building façades should be preserved. However, in practice, it is quite common that the façades suffer the deformations that bend the straight vertical lines. (2) *Projection of the spherical cameras*. The sensor acquiring the panoramas can be characterized by a spherical camera model, for which the images before mosaicking share the same projection center approximately. The pixels are warped from the unit sphere to a rectangular image using the equirectangular projection, which severely distorts the upper part of the panoramas.

Aiming at solving the above issues that impede the façades modeling from a single panorama, this paper proposes an approach using as-rigid-as-possible (ARAP) deformation and gnomonic rectification for the reparation of distorted panoramic images. Inspired by the prominent ARAP deformation of the surface meshes [17], we cover a gridded mesh on the panorama and cast the problem of straight line reparation to an ARAP-like deformation of the planar mesh [18], [19].



Fig. 1. Deformation and projection problems of panorama images. For the equirectangular projection, the top parts are shrinking and the vertical lines may be bended due to undesired deformations.

In addition, as the purpose is to straighten the bended façades structures, line constraints are also explicitly considered in the proposed methods. With regard to the anisometric problem caused by the equirectangular projection, this paper proposes a piece-wisely gnomonic rectification; with just a mild assumption that the façades are generally aligned vertically, the proposed methods find suitable tangent planes of the gnomonic projection and solve the metric problem without any additional position and attitude information of the sensor. The repaired and rectified images can be directly consumed by existing solutions as reference for 3D modeling.

In summary, the main contributions are two folds: (1) an improved ARAP-like deformation of the image with explicit line constraints and (2) a flexible tangent plane determination approach for the gnomonic projection of panoramas. The reminder of this paper is organized as follows. Section II presents the details of the above approaches. Section III demonstrates the experimental evaluations and conclusion remarks are given in the last.

II. METHODOLOGY

In order to increase the flexibility of the methods for the consumption by existing 3D modeling solutions [16], we confine the methods to a minimum amount of auxiliary information that is hardly collected or distributed. Therefore, this paper does not address the problems of systematic processing, calibration and integration of the mobile mapping system and only focuses on façades modeling from community street-view panoramas, e.g. the Google Street View [7]. An overview of the proposed methods is shown in Figure 2, which consist of ARAP deformation and gnomonic projection.

A. As-Rigid-As-Possible deformation for the reparation of bended structures

Inspired by ARAP deformation [17] of surface mesh, we would like to unbend the façades structures and meanwhile preserve the adjacent pixels as-rigid-as-possible. Therefore, we cover the original image with a gridded mesh [18] $\mathcal{M} = \{\mathcal{V}, \mathcal{F}\}$ on the image with $\mathcal{V} = \{\mathbf{v}_i(x_i, y_i)\}$ the vertices and $\mathcal{F} = \{f_i\}$ the quadrangle faces; in general, 40×20 grids are placed on each panorama. In order to repair the bended structures, line segments $\mathcal{L} = \{l_i(o_i, e_i)\}$ are detected using the popular line segment detector (LSD) [20], where o and e are the starting and end point of a segment, respectively. We are looking for the new positions of the all the above vertices, including \mathbf{v}' , \mathbf{o}' and \mathbf{e}' , under the energies defined below. The deformation

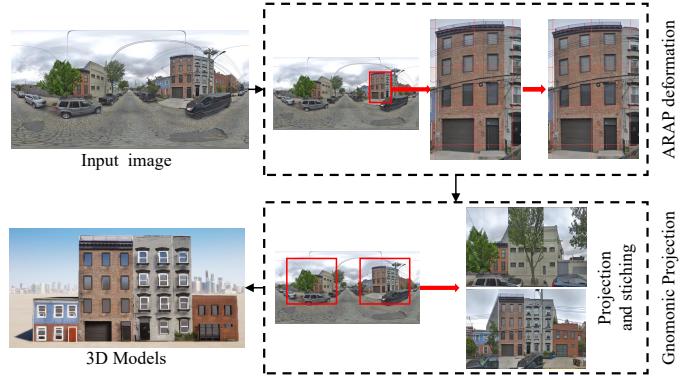


Fig. 2. Overview of the proposed methods. The ARAP deformation and gnomonic projection are responsible to repair the curved structures and to rectify the anisometric parts of the images, respectively.



Fig. 3. Illustration of the line constrained ARAP deformation.

of the image is cast to a series similarity transformations of the mesh vertices; the final image is the result of corresponding perspective transformations for each quadrangle using homographic transformation [21]. An illustration of the line constrained ARAP-deformation is shown in Figure 3.

a) *Mesh rigidity energy $E_r(\mathbf{v}')$* : The first term in the optimization is the mesh rigidity, taken from [18], which imposes a constraint on the quadrangle to preserve the original shape under a certain similar transformation SE(2), e.g. *scale*, *translation* and *rotation* $se(\mathbf{v}) = s(\begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \mathbf{v} + \begin{pmatrix} t_x \\ t_y \end{pmatrix})$; other deformation that leads to a non-square shape will be penalized. By substituting the four parameters with the least-squares solver [18], E_r could be written concisely as,

$$E_r(\mathbf{v}') = \frac{1}{|\mathcal{F}|} \sum_f \| (A_p(A_p^T A_p)^{-1} A_p^T - I) \mathbf{V}'_f \|^2, \quad (1)$$

where $A_p \in \mathbb{R}^{8 \times 4}$ and $\mathbf{V}'_f \in \mathbb{R}^8$ comprise the original and deformed coordinates of the four vertices enclosed by a face f , i.e. \mathbf{v} and \mathbf{v}' as below [18],

$$A_p = \begin{bmatrix} x_1 & -y_1 & 1 & 0 \\ y_1 & x_1 & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_4 & -y_4 & 1 & 0 \\ y_4 & x_4 & 0 & 1 \end{bmatrix}, \quad \mathbf{V}'_f = \begin{bmatrix} x'_1 \\ y'_1 \\ \vdots \\ x'_4 \\ y'_4 \end{bmatrix}. \quad (2)$$

b) *Vertical and parallel line rigidity energy $E_v(\mathbf{o}', \mathbf{e}')$ and $E_p(\mathbf{o}', \mathbf{e}')$* : For the vertical rigidity constraints $E_v(\mathbf{o}', \mathbf{e}')$, a subset of the detected or traced line segments $\mathcal{L}_v = \{l(\mathbf{o}, \mathbf{e})\}$ are selected interactively, which are placed to new positions. In addition, in order to keep the joint parts changing smoothly, we also optionally allow several endpoints $\mathcal{V}_e = \{\mathbf{e}_l\}$ to be kept fixed as the boundary constraints. Specifically for each vertical line segments l , we apply a transformation with scale

and rotation to the line and constrain it to be rigid with regard to the original one; because the rotation angle θ_l is directly computed from the misalignment to the vertical direction, the only additional unknown parameter is the scale s . Similar to Equation 1, by substituting the scaled line segments as $\tau = \mathbf{e} - \mathbf{o}$ and using $R_l = \begin{pmatrix} \cos \theta_l & -\sin \theta_l \\ \sin \theta_l & \cos \theta_l \end{pmatrix}$ to represent the known rotation matrix, we have an simplified rigidity constraints as Equation 3,

$$E_v(\mathbf{o}', \mathbf{e}') = \frac{1}{|\mathcal{L}_v|} \sum_l \| (R_l \tau (\tau^T \tau)^{-1} \tau^T R_l^T - I) \tau' \|^2. \quad (3)$$

For the parallel lines induced by the façade attachments, the target orientation is set to the average of all the parallel lines rather than vertical direction. In addition, we don't use fixed number of bins [19], [22], because horizontal parallel segment may have different orientations for the panoramas. The parallel rigidity energy E_p is defined similar to E_v , except for the matrix R is computed from the target orientation.

c) *Line coherent energy*, $E_c(\mathbf{v}', \mathbf{o}', \mathbf{e}')$: Until now, it should be noticed that the line segments $l(\mathbf{o}', \mathbf{v}')$ and the mesh grids \mathbf{v}' are still separated. Inspired by the rigid constraints of mesh grids that intend to preserve the shape of a quadrangle, we also keep the relative position of an endpoint fixed with respect to both the original and deformed grids. Specifically, the original endpoints \mathbf{o} and \mathbf{e} can be expressed by a bi-linear interpolation of the vertices of the grid, $\mathbf{o} = a_{11}\mathbf{v}_1 + a_{12}\mathbf{v}_2 + a_{21}\mathbf{v}_3 + a_{22}\mathbf{v}_4 = AV$; and the coefficient vector A is directly used for the deformed segments in $E_c(\mathbf{v}', \mathbf{o}', \mathbf{e}')$ by

$$E_c(\mathbf{v}', \mathbf{o}', \mathbf{e}') = \frac{1}{2|\mathcal{L}|} \sum_{o,e} \| (\mathbf{o}' - A_o \mathbf{V}_o') + (\mathbf{e}' - A_e \mathbf{V}_e') \|^2. \quad (4)$$

d) *Optimization*: By combining all the energies in Equations 1, 3 and 4, including the mesh rigidity E_r , vertical rigidity E_v , parallel rigidity E_p and coherent E_c , the total energy is a weighted summation among them; in addition, each term is normalized to the unit variance by applying *a priori* standard deviations before putting into the total energy as,

$$\min_{\mathbf{v}', \mathbf{o}', \mathbf{e}'} \lambda_r E_r + \lambda_v E_v + \lambda_p E_p + \lambda_c E_c, \quad (5)$$

where λ s are the weights. The above optimization is the standard least-squares problem and solved using the Ceres Solver from Google Inc. [23]. Specifically, the constraints of fixed endpoints in \mathcal{V}_e are explicitly constrained by setting the parameter blocks constant, which is supported by Ceres Solver. For the parameters λ , similar to previous work [19], we also award larger weights to the line structures; and $\lambda_r = 0.1, \lambda_v = 0.4, \lambda_p = 0.4, \lambda_c = 0.1$ are used in all the experiments below.

B. Gnomonic projection for panoramic images

Generally, the panorama is obtained by mapping multiple images acquired at the same time to a unit sphere and then unwrapped to a 2D image plane by equirectangular projection. Although the original panoramic image maintains the straight structures after the above ARAP deformation constrained by line segments, the images are still distorted caused by the



Fig. 4. Selection of the tangent points. (a) The tangent points located directly on the inflection points of the road marks and sidewalks and (b) when occluded, it may be also found from the inflection on the curves of vehicles and façades.

projection problem, especially for objects that are distant from the central horizontal line. Aiming at alleviating this issue, this paper proposes a piece-wise gnomonic projection method to correct the distortion.

For the equirectangular projection, a point in the object space is first projected onto the sphere with radius $R = W/2\pi$; and then the corresponding azimuth θ and zenith φ angles are directly used to unwrap the sphere onto a raster image with width of $W = 2\pi R$ and height of $H = \pi R$. Therefore, pixel coordinates (x, y) of the panoramas have explicitly one-on-one correspondences with the spherical coordinates (θ, φ) .

Rather than directly use the unscaled coordinates (θ, φ) for the raster unwrapping, gnomonic projection maps the sphere onto a tangent plane. The rationale behind the adopted approach for the rectification of façades area is that, as long as the selected tangent plane is approximately parallel to the façades, the distortion could be alleviated. Therefore, the key to a satisfactory projection is to find the good tangent points $T(\theta_0, \varphi_0)$. Then the projection is determined as,

$$\begin{aligned} y &= R \cot \varphi \\ x &= R \tan(\theta - \theta_0). \end{aligned} \quad (6)$$

Although no auxiliary navigation or mapping information is available, the determination of good tangent points is quite obvious in the original panoramas, as the inflection points of the boundaries for road features (Figure 4a), such as marks and sidewalks. Even if the features are occluded, the tangent point can also be approximated from the building façades (Figure 4b). In addition, we assume the sensor is leveled and the tangent points should lie in the center of a panorama, i.e. $\varphi_0 = 90^\circ$; therefore, only the horizontal coordinates of the tangent point are selected. If the moving direction of the vehicle that collects the panorama is parallel to the road, the two tangent points should have an interval of 180° in the spherical coordinate (or half width of the image); however, we don't adopt this assumption and always select two or even more tangent points in special scenarios, such as the crossroads.

Another practical issue for the gnomonic projection is the choice of receptive area on the tangent plane or the horizontal and vertical field of view. Because the areas exceeding $|\varphi - \varphi_0| = 90^\circ$ or $|\theta - \theta_0| = 90^\circ$ will project to a point at the infinite distance, the ranges should be chosen properly. In order to prevent unnecessary up-sampling and to avoid the ground areas, the range of φ and $\theta - \theta_0$ has to be limited. For instance, if $\varphi \in [15^\circ, 105^\circ]$ and $(\theta - \theta_0) \in [-60^\circ, 60^\circ]$ are selected



Fig. 5. ARAP deformations for both residential and commercial areas.

TABLE I

ACCURACIES OF THE DIAGONAL LENGTH (SHADED CELLS) AND CENTER OFFSET OF THE WINDOWS FOR A FIVE FLOOR BUILDINGS IN m .

	Column #1		Column #2		Column #3	
Floor #1	0.02	0.02	0.01	0.02	0.03	0.06
Floor #2	0.09	-0.01	0.13	0.08	0.09	0.01
Floor #3	0.18	-0.04	0.21	0.06	0.19	0.01
Floor #4	0.21	-0.04	0.19	0.05	0.2	-0.04
Floor #5	0.18	0.02	0.14	0.07	0.15	-0.02

and the distance from the sensor to the façade is 10 m , the horizontal and vertical receptive areas cover about 35 m and 37 m , respectively. In addition, a practical remedy to extend the horizontal areas is by mosaicking of adjacent images.

III. EXPERIMENTAL EVALUATIONS

A. Results of image deformation and rectification

For the experimental evaluations, we first evaluate the proposed methods with qualitative experiments on the line constrained ARAP deformation and rectification using the gnomonic projection. For the image deformation, the purpose is to unbend the façade structures. Shown in Figure 5 is two examples. The top and bottom rows indicate the original and the deformed panoramas, respectively. It could be noted that the differences between the two are quite subtle; this is expected because two out of a total of four energies in the ARAP deformation (Equation 5) are trying to keep the original shapes. However, in the enlarged cyan regions (right column of each area), we can clearly see that the oblique and curved façade structures have been straightened satisfactorily, with axis-aligned lines covered as references for better interpretation.

However, it is still difficult to use the deformed images in Figure 5 for 3D modeling, because the façades are not facing the view direction and severe anisometric distortion present in the regions with small zenith angle φ . Therefore, the rectification using the gnomonic projection with proper tangent points and FOVs are necessary. Figure 6 demonstrates the effects of the rectification of distortion and correction of the facing directions of the façades. After the rectification, the sliced images from different panoramas can also be mosaicked using the same scale settings. This can be achieved in most off-the-shelf 3D modeling solutions.

B. Results of 3D modeling

The deformed and rectified images are directly imported into [16] as the reference images for the 3D façade modeling



Fig. 6. Image rectification using the gnomonic projection for both the residential and commercial areas.

and the textured results are imported into the rendering engine for visualization [24]. The comparison between textured models using rectified and original images are demonstrated in Figure 7. It could be noted that even using a single panorama, it is still possible to create photorealistic and semantic-aware façade models. The LoD-3 attachments are reconstructed satisfactorily [1], such as the windows and doors. The environmental light reflections on the windows as shown in the right column of Figure 7 are essential for many applications requiring realistic rendering and can only be achieved with semantic-aware LoD-3 models.

In order to further evaluate the metric capability of the gnomonic projection, we also use the photogrammetric mesh models from aerial oblique images [5], which are available on the Google Earth, as reference to validate the geometric accuracies of reconstructed windows. The images are only scaled with the height of the buildings measured from the Google Earth and the photogrammetric meshes are used as reference to be registered in the local coordinate system. The center positions and the diagonal lengths of the windows of a five-floor building are compared. The results are shown in Table I with an RMSE for the diagonal lengths and center offsets at 0.15 m and 0.04 m , respectively, which indicates a satisfactory metric accuracy. In fact, the good alignment is expected, because the rectification has no distortion as long as all the objects are coplanar, which is generally tenable for windows on the same façades.

C. Discussion and limitations

Based on the experimental evaluations, we discuss some properties of the proposed method and indicate some possible limitations.

Flexibility: The proposed method only requires a single panorama, obtained from community distributors; auxiliary information may be helpful, but at the cost of the complexities in the system design, which leads to problems in the consumption by off-the-shelf 3D modeling solutions. In order to fulfil this requirement, the panorama has to be repaired with only image-space processing techniques, e.g. the line constrained ARAP deformation and the gnomonic projection with suitable tangent points.

Accuracy: As long as a single scale parameter is used, e.g. measured from known distance or Google Earth, we



Fig. 7. Comparison of the 3D textured models between the proposed methods and original panoramas. The cyan regions are enlarged in the right column to show the semantic information of the façades, e.g. windows and doors.

could achieve reasonable metric accuracies in a local Cartesian coordinate frame. This is because the assumption of coplanar façades structures are tenable in most cases and non-coplanar structures can be also extruded along the façade by human perception.

Limitations for high-rise buildings: However, the gnomonic projections have some limitations for high-rise buildings in metropolitan cities. As seen in Equation 6, the coordinate are defined in the tangent space, which is infinite when the zenith angle approaches zero. The high-rise buildings will cause problems in the top. Direct spherical projection may be a better approach.

IV. CONCLUSIONS

3D modeling of the building façades is an important topic toward fully exploitation of the LoD-3 reconstruction. This paper proposes a flexible and practical solution aiming to solve this issue from a single panorama. The core contributions underlying the solution consist of (1) a line constrained ARAP deformation that corrects the bended structures caused by sensor defects or orientation inaccuracies and (2) the rectification of distortion using gnomonic projection by proper tangent point estimation. The experimental evaluations have evidenced that the proposed image-space correction methods will balance well between the two competing desires, that want to keep the original images and fix the curved structures. In addition, the metric accuracy is also acceptable even only using the scale measured from a known distance. However, current solution still has some limitations and future research directions include automatic detection of the façade objects using deep learning [25] and alignment (or snapping) of the structured objects in a consistent manner [14].

REFERENCES

- [1] T. H. Kolbe, G. Gröger, and L. Plümer, “Citygml: Interoperable access to 3d city models,” in *Geo-information for disaster management*. Springer, 2005, pp. 883–899.
- [2] D. Duarte, F. Nex, N. Kerle, and G. Vosselman, “Damage detection on building façades using multi-temporal aerial oblique imagery,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2019.
- [3] T. Kelly, P. Guerrero, A. Steed, P. Wonka, and N. J. Mitra, “Frankengan: Guided detail synthesis for building mass-models using style-synchronized gans,” *arXiv preprint arXiv:1806.07179*, 2018.
- [4] P. Gurram, E. Saber, and H. Rhody, “A segment-based mesh design for building parallel-perspective stereo mosaics,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 3, pp. 1256–1269, 2009.
- [5] H. Hu, Q. Zhu, Z. Du, Y. Zhang, and Y. Ding, “Reliable spatial relationship constrained feature point matching of oblique aerial images,” *Photogrammetric Engineering & Remote Sensing*, vol. 81, no. 1, pp. 49–58, 2015.
- [6] F. Li, M. Lehtomäki, S. O. Elberink, G. Vosselman, A. Kukko, E. Puttonen, Y. Chen, and J. Hyppä, “Semantic segmentation of road furniture in mobile laser scanning data,” *ISPRS journal of photogrammetry and remote sensing*, vol. 154, pp. 98–113, 2019.
- [7] D. Anguelov, C. Dulong, D. Filip, C. Frueh, S. Lafon, R. Lyon, A. Ogale, L. Vincent, and J. Weaver, “Google street view: Capturing the world at street level,” *Computer*, vol. 43, no. 6, pp. 32–38, 2010.
- [8] B. Micusik and J. Kosecka, “Piecewise planar city 3d modeling from street view panoramic sequences,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 2906–2912.
- [9] S. Verstockt, M. Gerke, and N. Kerle, “Geolocalization of crowdsourced images for 3-d modeling of city points of interest,” *IEEE geoscience and remote sensing letters*, vol. 12, no. 8, pp. 1670–1674, 2015.
- [10] T. K. Dang, M. Worring, and T. D. Bui, “A semi-interactive panorama based 3d reconstruction framework for indoor scenes,” *Computer vision and image understanding*, vol. 115, no. 11, pp. 1516–1524, 2011.
- [11] W. Wahbeh and S. Nebiker, “Three dimensional reconstruction workflows for lost cultural heritage monuments exploiting public domain and professional photogrammetric imagery,” *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 4, 2017.
- [12] Y. Yang, S. Jin, R. Liu, S. Bing Kang, and J. Yu, “Automatic 3d indoor scene modeling from single panorama,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3926–3934.
- [13] Y. Li, Q. Hu, M. Wu, J. Liu, and X. Wu, “Extraction and simplification of building façade pieces from mobile laser scanner point clouds for 3d street view services,” *ISPRS International Journal of Geo-Information*, vol. 5, no. 12, p. 231, 2016.
- [14] S. Hensel, S. Goebbel, and M. Kada, “Facade reconstruction for textured lod2 citygml models based on deep learning and mixed integer linear programming,” *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 37–44, 2019.
- [15] Y. Jia, G. Cui, L. Kong, and X. Yang, “Multichannel and multiview imaging approach to building layout determination of through-wall radar,” *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 5, pp. 970–974, 2013.
- [16] SketchUp, “Sketchup: 3d design software — 3d modeling on the web,” <https://www.sketchup.com/>, 2019, accessed: August 6th.
- [17] O. Sorkine and M. Alexa, “As-rigid-as-possible surface modeling,” in *Symposium on Geometry processing*, vol. 4, 2007, pp. 109–116.
- [18] G.-X. Zhang, M.-M. Cheng, S.-M. Hu, and R. R. Martin, “A shape-preserving approach to image resizing,” in *Computer Graphics Forum*, vol. 28, no. 7. Wiley Online Library, 2009, pp. 1897–1906.
- [19] K. He, H. Chang, and J. Sun, “Rectangling panoramic images via warping,” *ACM Transactions on Graphics (TOG)*, vol. 32, no. 4, p. 79, 2013.
- [20] R. G. Von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, “Lsd: a line segment detector,” *Image Processing On Line*, vol. 2, pp. 35–55, 2012.
- [21] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.
- [22] C.-H. Chang, Y. Sato, and Y.-Y. Chuang, “Shape-preserving half-projective warps for image stitching,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3254–3261.
- [23] S. Agarwal, K. Mierle *et al.*, “Ceres solver,” ceres-solver.org, 2012, accessed: August 6th.
- [24] EpicGames, “Unreal engine: Twinmotion,” <https://www.unrealengine.com/twinmotion>, 2019, accessed: August 13th.
- [25] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” in *Advances in neural information processing systems*, 2015, pp. 91–99.