A SURVEY ON DESIGNING BUILDING MODELS FROM SINGLE IMAGE

Anagha S[1], Prof. Ajai Ram[2]

[1]*Student, Master of Technology, Computer Science & Engineering, RIT Engineering College Kottayam, Kerala, India*
[2]*Professor, Dept. of Computer Science and Engineering, RIT Engineering College Kottayam, Kerala, India*

---***---

Abstract - *Modeling buildings in 3D in real world have been useful for urban applications such as planning and visualization. Creating a virtual city is also demanded for computer games, movies etc. but it takes a lot of time to create numerous 3D building models. Various drawing tools like AutoCAD Software and other Texture mapping methods and Architectural modelling are present for creation of 3D procedural models. However, creating an architectural grammar can be difficult and time consuming. In this paper we are discussing some of image-based methods for creation of 3D models. Additionally, we are introducing a new approach is presented that allows users to automatically generate such a 3D building models from a single image of a building. The model allows to select photograph and the building silhouette is highlighted which will be the initial input to our method. In the first stage highlighted silhouette is unwrapped and its camera parameter are estimated and in next stage by combining estimated camera parameters and style parameters the building grammars are generated automatically. Finally, 2D shapes are designed to 3D building models. This methodology consists in simplifying the input image to match the visual appearance of synthetic training data, and in using optimization to refine the parameters estimated by CNNs. We can use this approach to generate a variety of 3D models of buildings from existing photographs.*

Key Words: **Architectural grammar, Texture Mapping, CNN, Procedural models**

1.INTRODUCTION

The interactive visualization of architecture and buildings provides a way to see current structures, as well as future tentative structures and changes to existing buildings. Virtual city creations can be useful in many fields like for computer game development, movies with digital content generation, urban planning etc. The needs for 3D city models are growing and expanding rapidly in various fields include urban planning and design, architecture, environmental visualization and many more[9]. In case of urban planning, modelling the appearance and behavior of urban spaces is a great challenge. An urban space is a complex collection of architectural structures arranged into buildings, parcels, blocks and neighborhoods interconnected by streets. Understanding, describing and predicting the appearance (e.g., creating 2D/3D geometric models) and behavior (e.g., simulating urban development over time) of cities is useful in a growing number of applications. Traditionally, modelling urban spaces[11] has been a rather manual task that consumes significant amounts of resources. With the growing requirements of quantity and quality in urban content, there is an imperative need for alternative solutions that allow for fast, semiautomatic urban modelling. Procedural modeling is a popular way to create virtual architectures because it can generate varying content through simple parameter changes. However, creating an architectural grammar can be difficult and time consuming. Generating virtual models of buildings from existing imagery can be approached from several directions (using machine learning like SVM, RF etc.). On one extreme, Multiview stereo focuses [4] on accurately reconstructing buildings from a multitude of images but requires many images or a carefully selected set. At the other extreme, previous single-image modeling methods yields similar-looking building models but require significant user effort or have very limited modeling abilities. Further, none of these methods yield semantic information or a parameterized procedural model as output.

We are proposing an interactive model that allows the user to generate such a grammar automatically using a single example photograph. The new approach described in this paper, combines procedural modeling with deep learning to automatically generate a parameterized 3D procedural model of a building from a single photograph. The user selects a photograph and highlights the silhouette of the target building as input to our method. Our pipeline automatically generates the 3D building models from building mass which is evaluated from silhouette image of building. We introduce a method to jointly estimate camera pose and building shape from a silhouette image. For combining each module deep learning is used.

2. RELATED WORKS

Generally, most of existing methods [12],[13],14],[15,[16],[17],[18] are focused of drawing architectures for 3D model generation and some others focused on creation of architectural grammars by using various methods. Gen Nishidar et.al [1] proposed a system in which automatically generates a procedural building model from a single photograph. Method make use of recognition CNNs to select appropriate procedural grammars and for parameter estimation CNNs to estimate their camera and building style parameters. System aims at capturing buildings overall shape, the layout of its facade and the style of its windows. To do so, system decompose the problem into logical stages (mass, facade, windows) and treat each stage with a common methodology that consists in simplifying the input to make it amenable to analysis by CNNs trained with synthetic data, and refining the output with custom optimizations. The resulting pipeline can generate a diversity of procedural buildings with no user interventions. Pascal Muller et.al [2] presents algorithms to automatically derive 3D models of high visual quality from single facade images of arbitrary resolutions. System combines the procedural modeling pipeline of shape grammars with image analysis to derive a meaningful hierarchical facade subdivision. The proposed method is designed to work with input textures of arbitrary resolution. The proposed solution consists of four parts organized as stages in a pipeline. This pipeline transforms a single image into a textured 3D model including the semantic structure as a shape tree. Method uses a top-down hierarchical subdivision analogous to splitting rules in procedural facade modeling.

Daniel G. Aliaga et.al [3] proposing inverse procedural modeling of buildings. An interactive system is developed in which user can draw the blocks and can give visualization. In this system grammar is constructed from photographed and subdivided buildings, enabling the rapid sketching of novel architectural structures in the style of the original. Using several captured models, here shows that novel buildings can be designed very quickly and rendered with realism or style comparable to the original structures. Also demonstrated that the extracted procedural rules can be easily adapted to interactive projective texture mapping and to no photorealistic rendering. Alex Colburn et.al [4] proposed image-based remodeling methodology that allows real-time photorealistic visualization. It demonstrated a new interface for home modeling and remodeling using a through-the-photograph methodology. A synthesis of computer vision technology, texture atlas construction, and user-interface results in a system that provides an intuitive way to realistically visualize remodeling ideas in real time as one edits the geometry. Interface supports the creation of concise, parameterized, and constrained geometry. Guillou Erwan et.al [5] introduced camera calibration and 3D reconstruction. The first one allows the calibration of a camera from a single image (a photograph or any kind of image) containing two vanishing points. The method determines the focal length, then the rotation and translation matrices describing the rigid motion between the world and camera coordinate systems. To compute the translation matrix precisely, system needs to know the length of a segment (or edge) in the real scene. After the calibration, reconstruction is affected by placing a rectangular box around each potential object; the faces of the box are assigned textures. The holes in the extracted textures corresponding to invisible parts of box faces are filled with other appropriate portions of the image. These textures depend on the single viewpoint. The same textures are used to render the 3D reconstructed model from a novel viewpoint. Haiyong Jiang et.al [6] introduced a new approach for procedural modeling in which expressive selections are used instead of the simple matching of labels in current grammars. It introduced a procedural modeling language to encode procedural objects using selection expressions that enables us to model with a global view of the data. In new system shapes are selected using selection-expressions instead of simple string matching used in current state-of-the-art grammars. A selection-expression specifies how to select a potentially complex subset of shapes from a shape hierarchy.

Ignacio Garcia-Dorado et.al [7] presented a procedural modeling sketching system that enables a novice user to create a grammar and a corresponding 3D geometry quickly and easily, without the need of writing procedural rules. Method achieves interactive sketching of complex grammars by both decomposing the creation process of hierarchical models into coarse-to-fine stages, and by leveraging modern machine-learning algorithms to recognize grammar snippets and estimate their parameters in less than a second. The user only performs interactive sketching and our system automatically generates the procedural model and its parameters yielding the sketched shape. At first the user sketches a few strokes of the current object type to specify the desired shape. Using a deep-learning based method, the system finds which pre-defined grammar snippets yield visually similar shapes and chooses the best one by default. The selected snippets and their parameter values are merged into a single grammar that represents the entire building. Suzi Kim et.al [8] proposed a method to generate a three-dimensional (3D) virtual model of an imaginary city from a single street-view image to represent the appearance of the city in a given input photograph. Proposed method generates a 3D city model, which represents the appearance of the street in a given street level photograph by guessing unknown factors affecting appearance of the city. The proposed method categorized the factors into city components and city properties, where the city components refer to the physical elements that make up a city, such as roads, vegetation, and water, and the city properties refers to intangible characteristics, such as street patterns, density, and building type. System makes use of GAN and CNN models to infer the unknown components and properties and the models are trained from the real data of existing cities. In the newly introduced system showed how segmentation affects perceived city properties and terrain and height maps through the proposed learning models, and generated 3D city models by plausibly combining the derived city properties on the height and terrain maps.

3. APPROACHES FOR DESIGNING BUILDING MODELS FROM SINGLE IMAGE

3.1 *3D Models Designing Using Photogrammetry Method*

In traditional photogrammetry, we usually use an image pair to reconstruct the 3D model of spatial object. But to some objects which have regular shapes such as building, because there are certain geometric relations among the out lines. In one suggested technique a single UAV image, at first, we get the inner elements (x_0, y_0) and f by checking up the camera, secondly, we divide the parallel lines in buildings into three groups which parallel X, Y and Z coordinate axis respectively and compute the three joint points, thirdly calculate the three angle elements A,α and κ to implement the image relative orientation, fourthly compute the scale between reality and image using the real coordinate information abstracted from 2D GIS databases, then on the basis of all the above parameters we can get the real height and the ortho-texture of buildings[10]. One another method with the help of the UMK 10/1318 photogrammetric camera about 25 images were taken in normal position and additional approximate 100 digital photos taken in digital camera [8]. For 3D generation, the characteristic points where digitized and contains in *.txt file which have 3D coordinates. File is imported to AutoCAD software for 3D generation.

The point cloud modeling: In the research paper of Street, Iasi [8] modeling the point cloud was performed by using lines and surfaces with the existing functions in the Cyclone v. 6.0 software. Modeling started with the segmentation procedure where the scanned points were approximately grouped according to the plan they belong. There were defined areas built by a closer point group which fit well to a plan and they define distinct elements of the facade. The surfaces created in Cyclone were exported as a *.dxf file which was imported into AutoCAD 2010 platform and after that the 3D model of the building was created and the distinct elements were organized on layers. To obtain a better representation, the textures were created using fragments from digital photos of the building facades.

3.2 *3D Model Designing Using Aerial Images*

Apart from the above stated method 3D model generation can be done using aerial image[8]. The main data source is aerial photographs from which contours and topography are detected which are used to generate orthophoto image and Digital Elevation Model (DEM). Orthophoto image are used to derive building footprints and DEM is file contains pixel locations recorded as rows and columns along with coordinate information. With this information the models are designed. Building texture editing to eliminate shadow effects and obstructing objects removal and mappings are done with the help of Adobe photoshop software.

Fig-1: Generation of 3d model using aerial image a) Aerial Image, b) 3d model generated

3.3 ADDITIONAL AND NEW APPROACH

The techniques discussed above involved additional software usage and also time consuming for large scale generations. A survey has been conducted based on these motivations which has reached in designing a new approach. In this approach user selected photograph of building which is the basic input to our system. From the image the silhouette of building gets captured and it is passed on to the parameter estimation phase. For each stage pipelined CNN is used which should be trained with corresponding datasets.

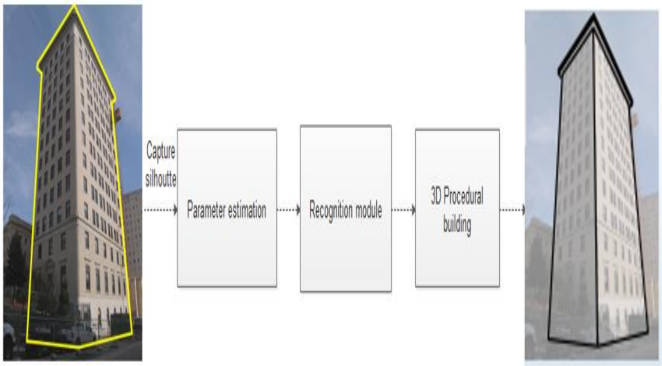

Fig -2: Different phases for designing 3D Model from image

The proposed model of Fig-2, has been decomposed mainly into three phases:

Parameter Estimation: In this phase captured building silhouette image is taken. From the building silhouette estimates the grammar parameters and calibrate the camera [1]. The CNN can jointly estimate camera and building parameters from the silhouette image. We are considering the intrinsic parameters of the camera (focal length and principal point) and the extrinsic parameters of the camera with respect to the building (rotation and translation). We cast camera intrinsic and extrinsic parameter estimation (8 parameters), as well as building mass parameter estimation (5-18 parameters). We adopt a camera model that covers at least most street-level photographs. The projection matrix can be calculated as:

$$M = K[R][T]$$

were K is the intrinsic camera matrix. R denotes the camera and rotation matrix and T represents the camera position(translation). As a result, we have a set of 8 parameters $\theta^C = (\varphi x, \varphi y, \varphi z, fov, Cx, Cy, tx, ty)$ to estimate, where $\varphi x, \varphi y, \varphi z$ denote the rotation angles around x, y and z axes and fov denotes the approximate horizontal field-of-view. Once the camera parameters are estimated, the focal length f can be easily calculated from fov. The captured silhouette is in non-regular shapes and hence model construction is tedious in that case. So, for solving this, after estimating these parameters the captured silhouette image is unwrapped or rescaled to a regular image.

Recognition Phase: In this phase the regular shaped silhouette images are taken as input. After unwrapping there occurs slight variations in building mass parameters and camera parameters. Again, these parameters are getting estimated in order, for the generation building mass grammars. Our approach estimates these parameters is inspired by the recent use of deep learning to predict viewpoint from a single image of a known object category and to predict parameters of procedural shapes drawn from a fixed viewpoint. By jointly estimates parameter values of the camera θ^C and building mass θ^M grammar for the building mass G^M is generated [1]. In our approach CNN is trained with building mass grammars in a systematic way and shapes are recognized and classified based on the results. Hence is this manner 2D shapes are obtained.

3D Procedural Building Phase: In this stage the 2D shapes so far generated is taken as input. This rendered 2D equivalents are converted to the 3D building models. For the projection transformation, in view plane want to establish the viewing-coordinate system, termed as view reference coordinate system. In the view plane a normal vector should be defined based on the building depth and horizontal field of view *fov* (field of view). Corresponding transformations and orthogonal parallel projections are given to the 3D models. As shown in figure, for the single image of the building by going through these stages final 3D model is generated.

4. CONCLUSION AND FUTURE WORKS

The generation of 3D building models have wide range of applications in various fields like urban planning, architecture, telecommunication, game development, planning and services etc. In this work we have reviewed on various techniques and methods used for creating 3D models. We have discussed about various architectural modelling methods in related work sections. Also, we have discussed various methods familiar in image-based 3D model generation, which comprised of model generation with photogrammetry and other is by using aerial images. In addition to that we have put forward a new approach for generating 3D models. Most approaches have been focused on the reconstruction of specific building models: rectilinear shapes and flat roofs [8]. In the new approach the drawback is it can design and generate building models exactly as in given image. The automation of extraction of buildings from various data sources is a very desirable and challenging yet equally difficult task because of the great number of possible building forms and shapes. As a future work research can be done for designing and developing new models from given image maintaining CNN accuracy.

REFERENCES

[1] Nishida, Gen, Adrien Bousseau, and Daniel G. Aliaga. "Procedural modeling of a building from a single image." Computer Graphics Forum. Vol. 37. No. 2. 2018.

[2] Müller, P., Zeng, G., Wonka, P. and Van Gool, L., 2007."Image-based procedural modeling of facades." ACM Trans. Graph., 26(3), p.85.

[3] Aliaga, Daniel G., Paul A. Rosen, and Daniel R. Bekins."Style grammars for interactive visualization of architecture.", IEEE transactions on visualization and computer graphics. 2007 Aug 20;13(4):786-97.

[4] Colburn, A., Agarwala, A., Hertzmann, A., Curless, B. and Cohen, M.F. "Image-based remodeling.", IEEE transactions on visualization and computer graphics.

[5] Guillou, E., Meneveaux, D., Maisel, E. and Bouatouch, K. "Using vanishing points for camera calibration and coarse 3D reconstruction from a single image.", The Visual Computer. 2000 Nov;16(7):396-410.

[6] Jiang. H, Yan, D.M., Zhang, X. and Wonka, P. "Selection expressions for procedural modeling.", IEEE transactions on visualization and computer graphics. 2018 Oct 23;26(4):1775-88.

[7] Nishida, G., Garcia-Dorado, I., Aliaga, D.G., Benes, B. and Bousseau, A "Interactive sketching of urban procedural models.", 2ACM Transactions on Graphics (TOG). 2016 Jul 11;35(4):1-1.

[8] Street, Iasi. "Comparative Study on Methods for 3D Modelling of Urban Areas." International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 39 (2012): B6.

[9] Kim, S., Kim, D. and Choi, S., "CityCraft: 3D virtual city creation from a single image.", The Visual Computer. 2020 May;36(5):911.

[10] Jizhou, Wang, L. Zongjian, and Li Chengming. "Reconstruction of buildings from a single UAV image."Proc. International Society for Photogrammetry and Remote Sensing Congress. 2004.

[11] Murphy, Maurice, Eugene Mcgovern, and Sara Pavia. "Historic building information modelling-adding intelligence to laser and image-based surveys." ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 3816 (2011): 1-7.

[12] Hoiem, D., Efros, A.A. and Hebert, M. "Automatic photo pop-up.", In ACM SIGGRAPH 2005 Papers 2005 Jul 1 (pp. 577-584).

[13] Sinha, S.N., Steedly, D., Szeliski, R., Agrawala, M. and Pollefeys, M. "Interactive 3D architectural modeling from unordered photo collections.", ACM Transactions on Graphics (TOG). 2008 Dec 1;27(5):1-0.

[14] Curless, Brian, and Marc Levoy. "A volumetric method for building complex models from range images." Proceedings of the 23rd annual conference on Computer graphics and interactive techniques. 1996.

[15] Kabolizade, Mostafa, Hamid Ebadi, and Ali Mohammadzadeh. "Design and implementation of an algorithm for automatic 3D reconstruction of building models using genetic algorithm." International Journal of Applied Earth Observation and Geoinformation 19 (2012): 104-114.

[16] Karsch, Kevin, Mani Golparvar-Fard, and David Forsyth. "ConstructAide: analyzing and visualizing construction sites through photographs and building models." ACM Transactions on Graphics (TOG) 33.6 (2014): 1-11.

[17] Yin, Xuetao, Peter Wonka, and Anshuman Razdan. "Generating 3d building models from architectural drawings: A survey." IEEE computer graphics and applications 29.1 (2008): 20-30.

[18] Yin, Xuetao, Peter Wonka, and Anshuman Razdan. "Generating 3d building models from architectural drawings: A survey." IEEE computer graphics and applications 29.1 (2008): 20-30.