

Some new research directions to explore in urban reconstruction

Florent Lafarge

INRIA Sophia Antipolis, France

email: firstname.lastname@inria.fr

Abstract—In this paper we present an update on the geometric modeling of urban scenes from physical measurements. This field of research has been studied for more than thirty years, but remains an important challenge in many scientific communities as photogrammetry, computer vision, robotics or computer graphics. After introducing the objectives and difficulties of urban reconstruction, we present a non-exhaustive overview of the approaches and trends that have inspired the research communities so far. We also propose some new research directions that might be worth investigating in the coming years.

I. INTRODUCTION

In recent years, the evolution of acquisition technologies and methods has translated into an increasing overlap of algorithms and data in computer vision, computer graphics, remote sensing and robotics communities. Beyond the rapid increase of resolution through technological advances in sensors, the line between laser scan data and photos is getting thinner. Combining, *eg* mobile laser scanners with panoramic cameras leads to massive 3D point sets with color attributes (on the order of 200M points per kilometer in a urban street). In addition, it is now possible to generate dense point sets not just from laser scanners but also from photogrammetry techniques when matching an acquisition protocol. Depth cameras are getting increasingly common and, beyond retrieving depth information, we can enrich the main acquisition systems with additional hardware to measure geometric information about the sensor and improve data registration.

These evolutions allow practitioners to measure urban environments at resolutions that were until now possible only at the scale of individual shapes. The related scientific challenge is, however, more than just dealing with massive data sets coming from an increase in resolution, as complex scenes are composed of multiple objects, each object being itself seen as an association of shapes. Understanding the principles that govern the organization of urban environments requires the analysis of structural relationships between objects and shapes.

The geometric modeling of urban scenes has received significant attention over time. This area of research, and especially the 3D reconstruction from physical measurements, is a topic of intellectual and commercial interest in many application domains. Computerized urban models are praised in urban planning for developing new plans in the context

of an existing environment, but also in navigation, digital mapping, electro-magnetic wave propagation study for wireless networks, emergency management, disaster control, mission preparation for defense, entertainment industry, etc.

Researchers have concentrated their efforts at three different scales: remotely sensed, terrestrial and indoor. The former has been deeply explored for several decades, mainly driven by the emergence of remote sensing in the eighties. At this scale, a description of the main urban objects is expected, for instance, building roofs and road networks. Terrestrial and indoor scales have been more recently addressed, driven by the advances of sensors in terms of quality and mobility. Embedded in cars or reduced to a simple web-cam, these sensors have provided new data measurements allowing the analysis of streets, facades or building rooms, for example.

II. SCIENTIFIC CHALLENGES

We can be distinguished three main challenges in the field of urban reconstruction: acquisition constraints, quality of models, and full automation.

Robustness to acquisition constraints. The acquisition process is usually a difficult task in the urban context. This produces defect-laden data. Noise is one of the typical defects. It can result from approximation in the data registration or directly from the sensor precision. Outliers are also frequent, especially with image stereo matching operations from textureless and reflective surfaces. Outliers also result from the presence of unwanted objects in scenes, for instance the temporary elements and the road signals when the modeling of buildings and facades is considered. Dealing with data which are heterogeneously sampled in the space is also a difficult problem. This arises, in particular, with laser sensors embedded into vehicles, the density of points decreasing according to the distance to objects. The most common and challenging defect remains the missing parts. Data can hardly cover entire complex environments because of the frequent occlusions. Geometric priors are typically exploited to explain such missing parts.

Quality of models. Depending on the application domain, different output properties can be expected. A result of *good* quality is not only a model with a high geometric accuracy, or faithfulness to the physical scene. A *good* computerized

representation can be also defined by (i) the model complexity measuring the degree of compaction of the output representation, (ii) the structural guarantees imposing global regularities on the geometry and semantics of the output, *eg* the respect to some level of details formalism (Fig. 1), and (iii) the visual aspect of the representation. The models must thus be measured and evaluated according to different criteria, usually in conflict between each other. Elaborating flexible metrics and strategies, which enable the combination of all these different criteria, constitutes one of the main challenges in the field. In particular, evaluation tools for measuring the geometric quality of models against Ground Truth are still largely undeveloped compared to other vision and graphics application fields. The recent benchmark proposed in [1] constitutes a preliminary step in this direction. The evaluation problem arises from the difficulties to (i) create accurate Ground Truth at the scale of a city, (ii) share non-public datasets, and (iii) propose relevant quantitative criteria that combined both geometric and semantic considerations. In our case, output models have been quantitatively evaluated from a restricted number of objects for which geometric and semantic information were available.

Full automation. One of the geometric modeling goals is to be as automatic as possible. Interactive modeling is usually recommended for architectural monuments and historical buildings, but remains ill-adapted to massive data for which considerable human resources would be required. Reaching full automation is an extremely difficult task as urban environments are complex and organized with a high degree of randomness, often resulting from an anarchical creation over time. Urban objects significantly differ in terms of diversity, complexity and density, even within a same scene. A predefined set of urban assumptions is rarely fully respected at the scale of a city. In practice, algorithms exploiting urban assumptions fails to model the entire scenes. To the contrary, algorithms omitting these assumptions are more flexible but the quality of models are usually lower. Faced with this dilemma, some scientists adopt an *automatic-then-interactive* strategy, the second step consisting in the interactive correction of the mistakes produced during the automatic step.

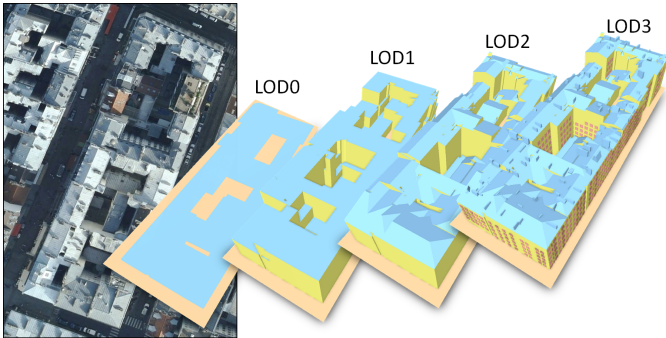


Fig. 1. Building reconstruction conforming to Levels Of Details used by cityGML [2]. LOD0 delineates the footprint of buildings. LOD1 represents the building volume with flat roofs. LOD2 provides additional details with piecewise-planar roofs. LOD3 provides further details such as roof super-structures, doors and windows.

III. APPROACHES AND TRENDS

The scientific literature related to the geometric modeling of urban environments is large and lies across several

communities such as computer vision, geometry processing, robotics and photogrammetry and remote sensing. The numerous approaches can be classified according to multiple criteria including acquisition specificities, characteristics of outputs, controllability, methodological foundations, and type of observed environments. Without being exhaustive, we present some existing works by listing several dualities. For a deeper review, the reader is invited to consult recent surveys [3], [4], [5], [6].

Reconstruction and generation. Reconstruction and generation of cities represents two distinct problems. Reconstruction is the process of creating a model as close as possible to data measurements in terms of accuracy [6]. Generation consists in artificially creating realistic models given some predetermined rules and procedural mechanisms [7], [4]. In recent years, these two distinct problems have tended to merge, in particular with recent works on inverse procedural modeling. This field of research constitutes one of the main challenges in city modeling.

Airborne and terrestrial acquisitions. These constitute the two main types of acquisitions for urban modeling problems. Terrestrial systems are suited to capturing vertical components such as facades. Data usually contain many occlusions as a scene is seen as a set of urban object layers from the sensors. Airborne and satellite systems allow the description of landscapes at bigger scales, and particularly the non-vertical components such as roofs or ground. Methods that exploit such data often assume a 2.5D representation of the scene in the sense that only one layer of objects is present.

Image and Laser. Geometric modeling of urban environments mainly relies on two types of measurements: Multi-View Stereo (MVS) imagery and Laser. As mentioned in [8], notable differences exist between these two inputs. Imagery has usually a better accessibility and coverage than Laser. Nevertheless, 3D information cannot be straightforwardly obtained from MVS images because camera calibration and image matching operations are required. These active research fields in computer vision have led to numerous surveys and benchmarks such as [9], [10] to cite just a few of them. To the contrary, Laser acquisition directly generates points in the three-dimensional space with high accuracy. The problem of recovering shapes and surfaces is, however, similar in both worlds, 3D-points being the reference element. Indeed, the use of points in MVS imagery as an intermediate step between images and surfaces is now a commonly accepted idea in the vision community, outclassing the direct use of implicit surfaces [11], [12]. In particular Multi-View Stereo systems are now mature enough to produce an accurate raw geometry from images, *eg* [13]. In many cases, images and Laser are combined to reinforce the modeling, *eg* [14]. In addition to MVS imagery and Laser, other types of acquisition are emerging, in particular, depth cameras such as Kinect enable the reconstruction of objects at short distances in real time [15].

Geometry and semantics. Geometry refers to questions of shape, size and relative position of objects in space. Semantics refers to the meaning and the nature of the objects composing a scene. Geometry and semantics are closely correlated in urban modeling. Semantics impacts on geometry in the sense that knowing the nature of an object allows us to adapt the modeling of the objects with specific geometric priors. For

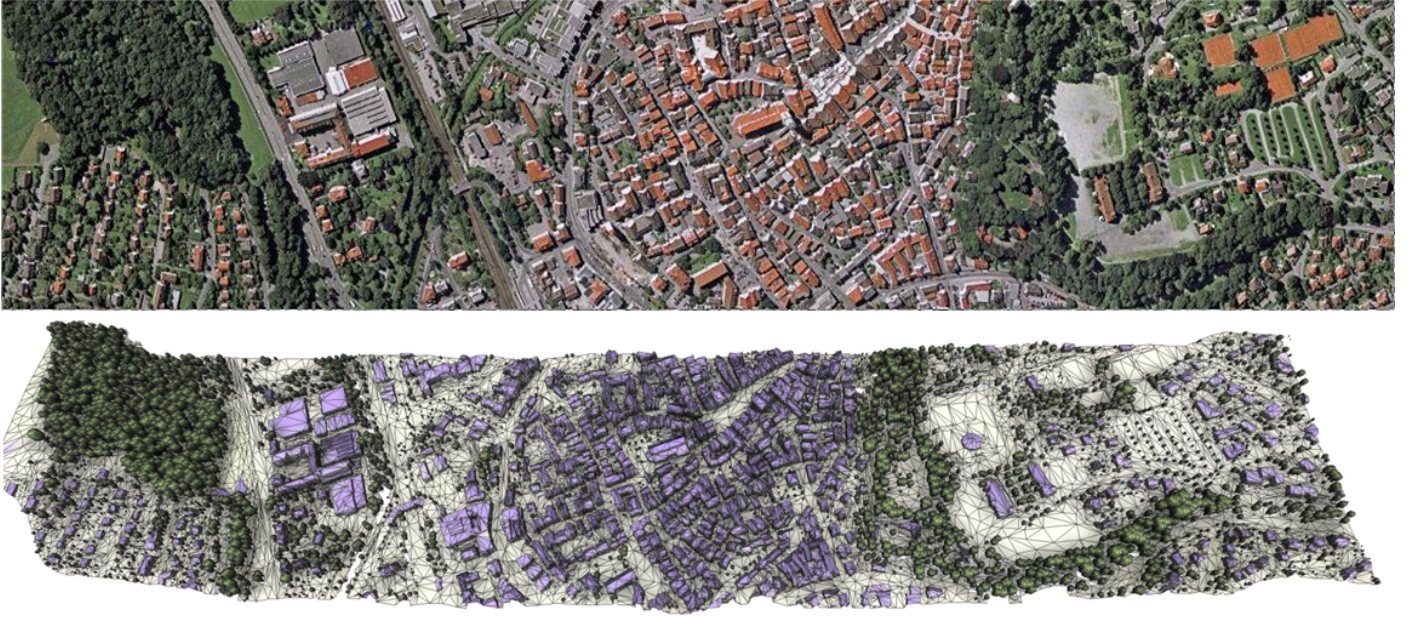


Fig. 2. Reconstruction of an urban scene (Biberach, Germany) from airborne Lidar scans. The output 3D model (bottom) is produced using a planimetric arrangement method [16]. Three types of urban objects are considered: buildings, trees and ground.

instance, piecewise-planar models are more suitable for modeling buildings that free-form surfaces which are more adapted for modeling trees. Geometry also impacts on semantics as the geometric relationships between objects can help to discover their nature. Recent works propose approaches combining geometry and semantics jointly, such as [17]. Most of existing approaches, *eg* [16] (see Fig. 2), rely on a *semantics-then-geometry* strategy, which might be less mathematically elegant than a joint analysis, but has demonstrated flexibility and scalability.

Free-form and structure. Objects can be modeled by free-form representations, or by more specific representations exploiting geometric primitives, and beyond them, structural relationships. Free-form representations have been deeply explored in literature, in particular for smooth shapes from nature and designers. Urban scenes are mainly composed of man-made objects for which the notion of structure is important. Structure is a generic term, not necessarily well-defined, that refers i) to the way the individual shapes are grouped to form objects, object classes or hierarchies, ii) to geometry when dealing with similarity, regularity, parallelism or symmetry, and iii) to domain-specific semantic considerations. Discovering structural relationships is of interest for i) consolidating and reinforcing the data, in particular in presence of occlusions and corrupted measurements, ii) increasing the geometric regularity of output models, and iii) simplifying the modeling with a solution space reduction [18]. Beyond conventional surface representations used in vision and geometry processing, the concept of hybrid surfaces collects the advantages of both free-form and primitive worlds, *ie* high robustness, low model complexity and structure-awareness. This concept has been applied to surface generation problems in different contexts, in particular, surface reconstruction and approximation from raw meshes [19] and from Multi-View Stereo images [20].

Local and global strategies. Contrary to local strategies, global strategies assume that entities composing a scene in-

teract, even if they are spatially far away from each other. The choice between local and global is usually a trade-off between output quality and performance. Local strategies are computationally less complex (and thus faster), whereas global approaches lead to output solutions with more regularities. The Markovian assumption constitutes an interesting alternative between purely-local and global strategies as it restricts the dependency of entities in a certain neighborhood. In particular, this assumption is suitable to parallelization strategies [21]. In geometric modeling, interactions between entities are usually spatial, and correspond to geometric constraints that can be either hard (binary condition) or soft (continuous score). Global strategies are commonly used for primitive-based surface reconstruction, constrained meshing and surface approximation.

IV. SOME NEW DIRECTIONS TO EXPLORE

Looking into the future, we believe some new research directions in the field will emerge in the coming years. These directions could have both a strong scientific impact for addressing the urban reconstruction challenges mentioned in Sec. II, and an important applicative impact in our everyday life with the creation of more intelligent city models.

Urban scale-space exploration. The structure within urban environments is not a fixed entity, but it evolves depending on the scale at which the scene is analyzed. Beyond the detection of geometric primitives and the discovery of structural relationships, it seems crucial to design algorithms that provides control upon the structure. In particular, one of the main challenges is to explore the solution space across scales and automatically select the structure, for instance as the best trade-off between complexity and faithfulness to input data, where complexity relates to the enumeration of structural rules and their parameters.

Physical coherence. To constrain the solution space of

ill-posed urban modeling problems, one interesting research direction is to take into account physical principles in addition to geometric, semantic or structural considerations. Exploiting physical coherence is not only a means for reinforcing the method efficiency; it is also a goal for producing 3D models that conform to physical principles such as self-supporting masonry structures, or constraints related to manufacturing and 3D-printing. In free-form architecture modeling, some recent works explore this research direction, for instance for tiling a surface with specific manufacturing and machining constraints.

Functionality. Beyond modeling of objects and scenes as sets of structured objects, discovering their function is another important challenge. Form follows function is a common principle in design and architecture: the shape of an object should be primarily based upon its intended function. Geometry, structure, semantic and physical coherence contribute to characterizing the nature of objects, and can be further exploited to understand their utility and to specialize their computerized modeling. For urban modeling, one objective is to understand the function of a building by analyzing these different characteristics. Some preliminary works have been proposed at the scale of individual objects [22], and it is still a scientific challenge to extend some of these ideas to large scale scenes.

Community data. During the last decades, geometric modeling issues on urban environments have been largely tackled from specialized sensors as airborne/satellite stereoscopic imagery or Laser scanning. Today, this data acquisition paradigm is completely reassessed with the emergence of new acquisition procedures that allow non-specialized people to freely access and enrich big datasets. An increasing variety of sensors is progressively disseminated everywhere; the best illustration is probably the 1.5 billion smartphones interacting in the world. The emergence of such "community data" coupled with the expanding computational resources constitute a great opportunity to propose efficient solutions to two big recurrent limitations in urban modeling: the low coverage of the specialized data (only a hundred cities are digitalized in 3D in the world) and the lack of flexibility of existing methods that are designed from a specific type of data to produce standardized 3D models. This new paradigm leads to an entire rethinking of the existing algorithm designs towards more flexibility and different data specificities, going from specialized, rare, private, expensive and accurate to multi-sourced, massive, public, free and defect-laden measurements.

Dynamical urban environments. The fact urban environments permanently evolve in time provokes considerable efforts for detecting changes and updating models. Beyond change detection and city model updating, which are both traditional problems largely addressed in the literature, one major scientific challenge is to understand tendencies and even anticipate the evolution of urban environments in terms of geometry. Recent works in vision [23] have demonstrated that prediction functions can be efficiently designed to forecast human actions from image sequences. In urban modeling, such prediction functions could be created by analyzing geometric variations along time from a big flux of information, typically community data. Many urban indicators could be studied, going from the expansion/shrinking directions of cities to evolution of architectural style through to the road network

complexity.

REFERENCES

- [1] F. Rottensteiner, G. Sohn, J. Jung, M. Gerke, C. Baillard, S. Benitez, and U. Breitkopf, "The ISPRS benchmark on urban object classification and 3d building reconstruction," in *Proc. of the ISPRS congress*, 2012.
- [2] G. Groger and L. Plumer, "Citygml interoperable semantic 3d city models," *Journal of Photogrammetry and Remote Sensing*, vol. 71, 2012.
- [3] H. Mayer, "Object extraction in photogrammetric computer vision," *Journal of Photogrammetry and Remote Sensing*, vol. 63, no. 2, 2008.
- [4] C. Vanegas, D. Aliaga, P. Mueller, P. Waddell, B. Watson, and P. Wonka, "Modeling the appearance and behavior of urban spaces," in *EUROGRAPHICS State of the Art Reports*, 2009.
- [5] N. Haala and M. Kada, "An update on automatic 3d building reconstruction," *Journal of Photogrammetry and Remote Sensing*, vol. 65, no. 6, 2010.
- [6] P. Musialski, P. Wonka, D. Aliaga, M. Wimmer, L. Van Gool, and W. Purgathofer, "A survey of urban reconstruction," *Computer Graphics Forum*, vol. 32, no. 6, 2013.
- [7] P. Muller, P. Wonka, S. Haegler, A. Ulmer, and L. Van Gool, "Procedural modeling of buildings," in *SIGGRAPH*, 2006.
- [8] F. Leberl, A. Irschara, T. Pock, P. Meixner, M. Gruber, S. Scholz, and A. Wiechert, "Point clouds: Lidar versus 3d vision," *Photogrammetric Engineering and Remote Sensing*, vol. 76, no. 10, 2010.
- [9] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in Computational Stereo," *Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 8, 2003.
- [10] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A comparison and evaluation of multi-view stereo reconstruction algorithms," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2006.
- [11] H. Vu, R. Keriven, P. Labatut, and J. Pons, "Towards high-resolution large-scale multiview," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- [12] J.-M. Frahm, P. Fite-Georgel, D. Gallup, T. Johnson, R. Raguram, C. Wu, Y.-H. Jen, E. Dunn, B. Clipp, S. Lazebnik, and M. Pollefeys, "Building Rome on a cloudless day," in *ECCV*, 2010.
- [13] Acute3D, <http://www.acute3d.com/>, 2014.
- [14] C. Frueh and A. Zakhor, "Constructing 3D City Models by Merging Ground-Based and Airborne Views," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2003.
- [15] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon, "Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera," in *ACM symposium on User interface software and technology*, 2011.
- [16] F. Lafarge and C. Mallet, "Building large urban environments from unstructured point data," in *IEEE International Conference on Computer Vision*, 2011.
- [17] H. Lin, J. Gao, Y. Zhou, G. Lu, M. Ye, C. Zhang, L. Liu, and R. Yang, "Semantic decomposition and reconstruction of residential scenes from lidar data," *ACM Tran. on Graphics*, 2013.
- [18] J. M. Coughlan and A. L. Yuille, "The Manhattan world assumption: Regularities in scene statistics which enable Bayesian inference," in *Proc. of Neural Information Processing Systems*, 2000.
- [19] F. Lafarge, R. Keriven, and M. Bredif, "Insertion of 3D-Primitives in Mesh-Based Representations: Towards Compact Models Preserving the Details," *IEEE Trans. on Image Processing*, vol. 19, no. 7, 2010.
- [20] F. Lafarge, R. Keriven, M. Bredif, and H. H. Vu, "A hybrid multi-view stereo algorithm for modeling urban scenes," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, 2013.
- [21] Y. Verdie and F. Lafarge, "Efficient Monte Carlo sampler for detecting parametric objects in large scenes," in *European Conference on Computer Vision*, 2012.
- [22] H. Grabner, J. Gall, and L. Van Gool, "What makes a chair a chair?" in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011.
- [23] K. Kitani, B. Ziebart, J. Bagnell, and M. Hebert, "Activity forecasting," in *European Conference on Computer Vision*, 2012.