17^{th} *Computer Vision Winter Workshop*
Matej Kristan, Rok Mandeljc, Luka Čehovin (eds.)
Mala Nedelja, Slovenia, February 1-3, 2012

Rapid 3D City Model Approximation from Publicly Available Geographic Data Sources and Georeferenced Aerial Images

Markus Rumpler, Arnold Irschara, Andreas Wendel and Horst Bischof
Institute for Computer Graphics and Vision
Graz University of Technology, Austria
{rumpler, irschara, wendel, bischof}@icg.tugraz.at

Abstract. *We present a fully automatic, object-centered approach for creating 3D city models by combining public domain geographic information with highly overlapping aerial images. In a first step, we create a polygonal 3D surface mesh from terrain elevation data provided by the Shuttle Radar Topography Mission. Second, we transfer building outline polygons from 2D vector maps provided by the OpenStreetMap project to georeferenced aerial images and extract image patches of individual buildings. Subsequently, we perform dense multi-view image matching to estimate a depth map for the roof pixels from which we approximate building heights and roof structures. This allows to reconstruct a polyhedral 3D city model of an area of 1 km² with more than 700 buildings in less than 10 minutes on commodity hardware. Our model satisfies the CityGML Level-of-Detail accuracy requirements of +/– 2 m, evaluated by comparison to reference height data from airborne laser scanning. The obtained textured meshes are suitable for city planning, interactive visualization or simulations and may serve as base meshes for future refinement of the geometry.*

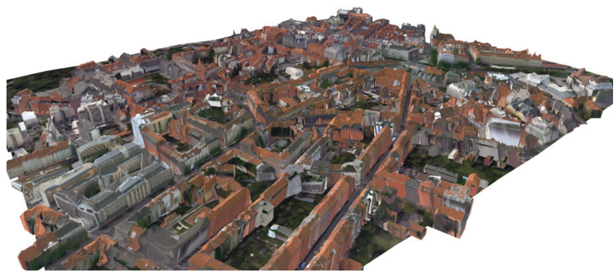


Figure 1. Visualization of a textured 3D city model created by our method, using public data and low-resolution aerial images in less than 10 minutes on commodity hardware.

1. Introduction

Virtual 3D city models started to evolve during the mid-1990s, when providers of geographic information systems (GIS) got interested in 3D versions of traditionally 2D geospatial data [13]. Simple 3D city models for visualization purpose were generated from textured digital surface models (DSM) of urban landscapes using semi-manual photogrammetric workflows.

While applications like wireless signal propagation or simulations for natural disaster management work with fairly generalized block models of buildings placed on top of a digital terrain model (DTM), others such as true-orthophoto generation need more complex and complete geometries. The concept of the *virtual habitat* [13] describes the most sophisticated form of semantically interpreted, photorealistic urban models, where each building, street sign, water body and tree exists as an individual, separate object.

A common information model named CityGML [11] has been proposed to distinguish between different representations of 3D urban objects. The standard defines five levels of detail (LODs) ranging from the bare Earth's surface up to the most advanced level with modeled building interiors, proposing accuracy requirements for each of these levels. In this work we consider LOD 1 and 2, which corresponds to models with flat roofs and models with approximated building heights and basic roof shapes, respectively.

Existing methods to recover basic roof shapes (LOD 2) from elevation data are either based on parametric shape fitting [2], rely on segmentation or feature recognition [19] or perform DSM simplification [21]. Many approaches utilize multiple data sources for roof shape reconstruction, e.g. LiDAR combined with 2D map data [2] or aerial stereo images [9]. A good overview on 3D building reconstruction can be found in Haala *et al*. [8].

In contrast, the OSM-3D[1] project [14] uses only publicly available elevation data and community generated 2D vector maps to create an interactive three-dimensional polygonal mesh-based map visualization. Polygonal meshes have several advantages: They allow a compact, memory efficient representation of irregular structure and varying complexity, simple distribution and visualization. It shows that with textured meshes, even relatively simple building models and coarse roof geometry is sufficient to generate pleasing visualizations. However, a drawback of OSM-3D is that the Level-of-Detail in which the buildings are modeled is solely based on user annotations, but height and roof shape attributes are not comprehensively available in the 2D map data.

In this work, we propose an approach to combine publicly available elevation and 2D map information with multi-view image matching on low-resolution aerial images for the reconstruction of roof structure and building height approximation. The fusion of existing geometric and semantic prior information allows to rapidly generate coarse polyhedral 3D meshes for city models of LOD 2 (Figure 1).

Our algorithm favors an object or world centered approach rather than an image based one. The obtained models can serve as base meshes for future refinement of the geometry without the need for fusion of multiple depth maps. If city models with enhanced details are required, information from full-resolution aerial image matching may be integrated and one can take advantage of the known (approximate) geometry of the scene by performing visibility checks and occlusion handling.

2. Data Sources

In our approach, we combine multiple data sources to reconstruct a 3D city model. We use existing information for buildings from freely available 2D map data and combine them with multiple overlapping aerial images, from which we can reconstruct the individual building heights and roof structures. In order to create geographically and visually realistic 3D city models, we need information about the terrain height in addition to map data, which we obtain from a publicly available digital elevation model (DEM). We use the terms DTM and DEM synonymously to describe the bare Earth's surface in contrast to DSMs including all objects on it.

2.1. Terrain Elevation

The Shuttle Radar Topography Mission (SRTM) [4], obtained a DEM of the Earth at near-global scale, covering about 80% of the Earth's total landmass. SRTM produced the most complete and high-resolution digital topographic database of the Earth.

The dataset is available at 1 arc-second resolution (SRTM-1, approximately 30 meters) over the United States and its territories and at 3 arc-second resolution (SRTM-3, approximately 90 meters) for the rest of the world. The international SRTM-3 elevation data is available in 2.5D raster format with 1201 columns and 1201 rows of height samples, meaning elevation measured in meters above sea level. Each SRTM data tile covers 1 by 1 degree of latitude and longitude.

2.2. OpenStreetMap 2D Vector Data

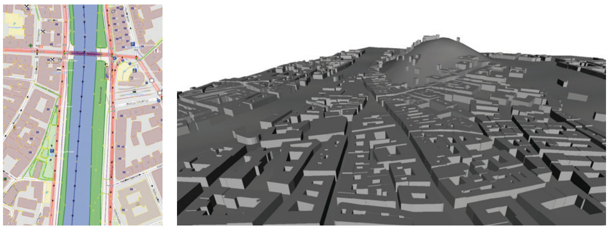
The way how geographic information is being collected, processed and distributed has changed in many ways during recent years. While these tasks were formerly more centralized and dedicated primarily to professionals in the field of surveying, geodesy or cartography, there is now a shift towards publicly and freely available geographic information systems (GIS) maintained from both commercial vendors and volunteers all over the world [5, 6].

Especially collaborative online community projects such as OpenStreetMap[2] (OSM) lead to a massive increase of data provided by citizens in a crowd sourcing manner, significantly promoting the amount, availability and nature of geographic information. For this reason, OSM content is a great source for geographic information to enrich and assist the creation of virtual city models.

Besides street networks and manifold points of interest (POIs), the OSM project provides outlines of buildings for many cities around the world (Figure 2). The maps are made from only a few simple elements (i.e. data primitives), namely nodes, ways and relations to model the geometry of objects as 2D vector data. An arbitrary number of so called tags can be assigned to an element to describe its properties or meaning (i.e. the semantics of an object) and define logical things like streets, POIs or buildings.

[1] `http://www.osm-3d.org/`

[2] `http://www.openstreetmap.org/`



(a) (b)

Figure 2. (a) OpenStreetMap 2D vector data showing streets, building outlines, paths of rivers and points of interest (POIs). (b) City model generated from OpenStreetMap and terrain elevation data corresponding to LOD 1.

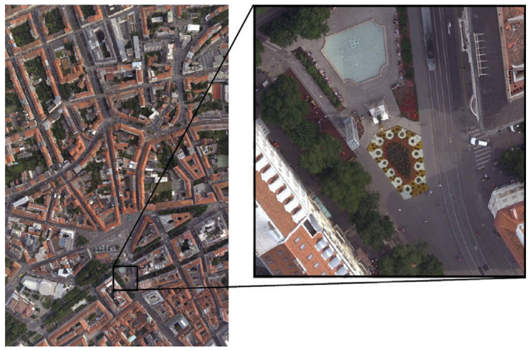


Figure 3. High resolution aerial image comprising 7500×11500 pixels at a ground sampling distance (GSD) of 8 cm/pixel, captured by the Microsoft Vexcel UltraCamD.

The OSM map data can be downloaded in XML format, containing blocks of nodes, ways and relations. Nodes are the basic elements. Each node stores a single geospatial point (longitude and latitude). A way is an ordered interconnection of at least two nodes that share a common property and is used to describe polygonal objects such as streets or building outlines. Atrium house footprints for example are modeled as multi-polygon relations.

2.3. Aerial Images

In the last two decades digital cameras started to dominate the market for aerial photogrammetry and photogrammetric DSM generation. In this time, the cameras reached a mature technical state and hence provide the necessary geometric and radiometric stability and resolution. Together with state-of-the-art multi-image matching techniques, aerial cameras compete with the legacy of airborne LiDAR systems, those were for a long time the preferred tool for collecting high quality elevation data [7, 12]. In addition, the cameras feature the advantage that the acquired digital images are readily suitable for automatic processing by computers. Figure 3 illustrates a typical high resolution aerial image captured during a mapping survey.

Today, typical airborne photogrammetric surveys are flown with at least 80% forward overlap and 60% side-lap. At flying height of 1000 m, a ground sampling distance (GSD) of 8-10 cm/pixel is obtained. The high resolution imagery and high redundancy opens the possibility to generate detailed maps of the environment using state-of-the-art Structure-from-Motion (SfM) techniques [10].

3. 3D City Model Generation

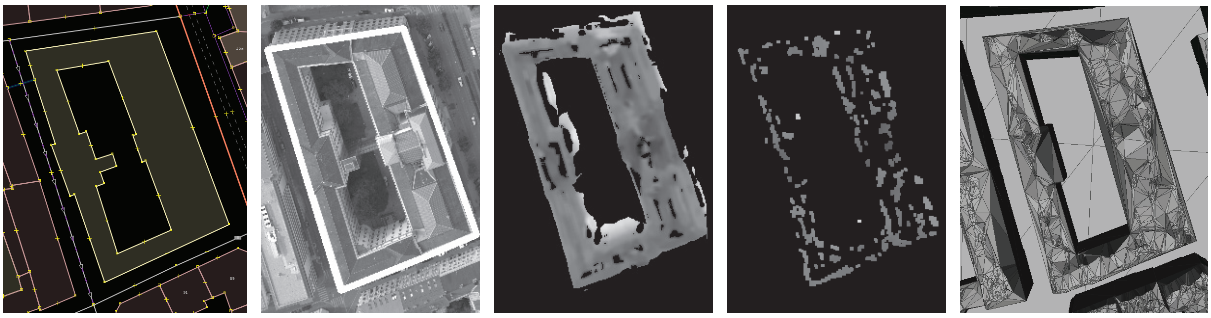
Our method fuses multiple data sources to create 3D city models as a three dimensional surface mesh. SRTM height information and OSM map data comes in the "geographic" WGS84 GPS coordinate frame, where the oriented camera block may be given in an arbitrary local Euclidean coordinate system. The camera block is georeferenced with an appropriate transformation into the global coordinate system which for example can be determined using a similarity transform between camera centers in the local frame and known real world positions of the cameras.

We need to transform DEM, OSM data and the oriented georeferenced aerial camera views into a common coordinate frame. In our system, we convert WGS84 ellipsoidal coordinates into a metric Euclidean East-North-Up (ENU) UTM projection.

The workflow of city model construction incorporates terrain model generation, building reconstruction for height and roof structure approximation, modeling of polyhedral building objects and texturing of the model. Mesh and point cloud processing is performed using the Computational Geometry Algorithms Library (CGAL) [1].

3.1. Creating a Terrain Mesh

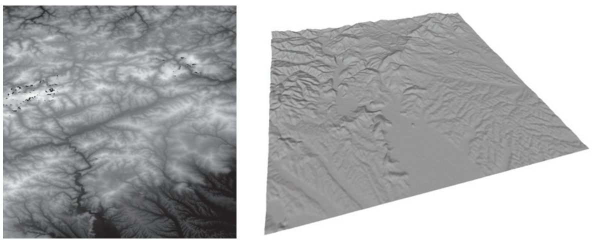
We start our 3D city modeling approach by generating a digital terrain model from the SRTM terrain height information. We create a 3D surface mesh (Figure 5) from the 2.5D raster format elevation data to represent geometry as a triangulated irregular network (TIN) [15]. To construct a triangle mesh from the unconnected sample points, we perform a 2D Delaunay triangulation and afterwards assign the corresponding height to each inserted vertex. Data voids



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

Figure 4. Building reconstruction using OpenStreetMap building footprints and aerial images. (a) Building polygon from OSM map data in vector format. (b) Extracted key view image patch with overlaid building polygon at approximated building height. (c) Depth map of building roof structure, depth values with large image correlation scores are removed and the result is smoothed for outlier suppression. (d) Simplified point cloud for data reduction (sparse points dilated for better visibility). (e) Reconstructed polyhedral building model with meshed roof points.

present in the raw SRTM data due to shadowing, topographic variations or other radar-specific causes need no further processing, as the missing areas are implicitly interpolated through the mesh triangles of adjacent valid data points.



(a) (b)

Figure 5. Raw SRTM elevation data and 3D surface mesh rendering. (a) Terrain elevation data in raster format. (b) Earth's surface represented as a 3D triangle mesh, which corresponds to CityGML LOD 0.

3.2. Reconstruction of Building Models

Our approach for 3D building model reconstruction is based on three main steps to produce polyhedral building objects: (i) building outline extraction from OSM map data and view selection, (ii) dense depth estimation and (iii) robust height approximation and 3D roof structure recovery.

3.2.1 From Footprints to Image Patches

For each building polygon extracted from the OSM 2D vector data (Figure 4(a)), we calculate its projection onto the terrain model and interpolate a ground height of terrain mesh points. Then, we back-project its bounding box and center point into all cameras to determine aerial views in which the current building is visible. Based on the viewing angle, we decide which of the selected views is best suited to serve as key view for multi-view building reconstruction. Key views are preferred to be nadir views, or the most orthogonal view with respect to the Earth's surface.

Subsequently, we extract an image patch from the key view (Figure 4(b)). Since we are satisfied with a fast approximation of the building height and a coarsely reconstructed roof structure, we do not need to process the full-resolution aerial images. Hence we restrict the maximum building patch size for multi-view matching and scale the input images accordingly.

3.2.2 Dense Multi-View Image Matching

The multi-view dense matching algorithm is based on the plane sweep principle [3], as illustrated in Figure 6, that enables an elegant way for image based multi-view reconstruction since image rectification is not required. The method allows reconstruction of depth maps from arbitrary collections of images and implicit aggregation of multiple view's matching costs. The redundancy from multiple views hereby contributes to scene completeness by capturing otherwise occluded areas, improves depth accuracy and increases robustness against mismatches in the presence of noise [18].

With the plane sweep method, the scene is iteratively traversed by parallel planes aligned with a reference view and positioned at an arbitrary number of discrete depths. Each sensor view i is then projected onto the current 3D key plane $\Pi_d = (n^\top, d)$ at depth d according to the epipolar geometry. The mapping

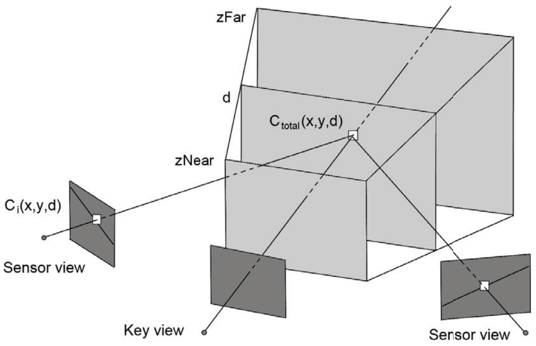


Figure 6. The scene is traversed by planes parallel to the key view. For each discrete depth, sensor images are projected onto the plane and similarity compared between pixels of key and sensor views by calculating cost values $C_i(x, y, d)$. A cost volume is filled with the accumulated matching scores. A final depth value for each pixel is extracted by a winner-takes-all (WTA) strategy.

is given by a homography $H(d)$ which can be computed directly from the camera matrices K and K', the relative rotation R and translation t between key and sensor camera, the normal vector $n^\top$ of the plane and the current depth d as

$$H_i(d) = K_i' \left(R_i - \frac{t_i n^\top}{d} \right) K. \quad (1)$$

Once we have found an appropriate camera to serve as the reference camera and an already extracted building patch, we select appropriate neighbor views in which the building is visible by back-projection of the building footprint into all formerly selected potential neighbors. In addition to the input images and cameras, the algorithm needs information about the geometric extent of the reconstructed scene as initialization. We need to define the maximum range or interval in which possible depth values can occur.

The plane sweep technique is based on the idea that, if the plane at a certain depth passes exactly through the object's surface we want to reconstruct, then, under constant brightness conditions, the appearance of the image points (i.e. the color or intensity values) in the key view should match with those of the projected sensor image. For that reason we need to determine a matching score between key and transformed sensor images to measure pixel similarity. We calculate the zero-mean normalized cross correlation (ZNCC) as cost function and accumulate the costs from all sensors to fill a 3D cost volume for all depth hypotheses.

We extract final depth values from the volume by a simple winner-takes-all (WTA) [20] strategy in order to achieve high performance and low memory requirements. We select the depth where the accumulated cost has its minimum along the line of sight through the volume, as we assume the correct depth as the one with the lowest cost value assigned. Optionally, fast cost volume filtering [17] can be applied for considering local neighborhood information for depth extraction. Figure 4(c) shows a resulting WTA depth map of a roof structure for one single building.

3.2.3 Approximating Building Height and Roof Structure

Given the estimated depth maps, we are ready to approximate building heights and generate meshes of the building's roof structure. To achieve this goal, several steps are necessary. Due to matching ambiguities and the WTA approach, resulting depth maps are noisy and contain outliers which are filtered by applying Gaussian smoothing.

Since the depth maps may contain depth estimates from close-by trees or parts of surrounding buildings, we need to mask out clutter which does not belong to the building. Furthermore, depth maps may be computed from key views imaging the building from oblique angles, instead of perfect nadir views. Then depth maps are likely to show parts that do not belong solely to the roof of the building, but also depth values belonging to the facades. We circumvent this by an orthographic projection of the depth map onto the X-Y plane of the coordinate system and masking out depth values which are not inside the building polygon.

To turn the roof structures into a compact mesh representation, the resulting point clouds corresponding to the depth maps need simplification to reduce the amount of data. We apply CGAL's point set processing features [1] for outlier removal and apply clustering based simplification. The algorithm sorts the input points according to the squared distance to the k nearest neighbors and removes a user-specified fraction of outliers. Afterwards, the point set is simplified in a way that, based on a regular grid, points sharing the same grid cell are clustered by picking one representative either randomly or by averaging all points in a cluster (Figure 4(d)).

We calculate the median depth from the reduced sparse roof points to robustly estimate a median height for the building. This height is used to extrude the building walls from the footprint points.

The 3D surface mesh for the roof is generated in the same way as the mesh generation of the DTM. A triangulation of the remaining sparse point cloud is obtained by 2D Delaunay triangulation with additional constrained edges at the building outlines, followed by height value assignment for each vertex (Figure 4(e)). Depending on the desired level of detail polyhedral building models can be modeled either with flat roofs or with approximated roof structure.

3.3. Texturing

We employ the aerial images as color textures to make the city model look more realistic. We perform multi-image mesh texturing, taking into account all views used for matching. For each triangle face of the model, we first determine all aerial cameras from which it is visible. This is the case if the triangle's face normal points towards the camera's view vector. If only one view should be used for texturing, we decide for the "best" one, that maximizes the size of the triangle's back-projection into the image in terms of the pixel area. In other words, we select the view, which is the one that sees the triangle face most closely from an orthogonal, perpendicular direction, minimizing the angle difference between face normal and camera vector.

The described approach works quite well in practice but can lead to visual artifacts if adjacent triangles are textured from different images with significant illumination changes. Simple averaging or a more sophisticated method for texture blending may be applied.

4. Results and Discussion

We use the following parameters to run our algorithm. From OpenStreetMap, we extracted all ways and relations tagged either as "building" or "multipolygon". The maximum image patch resolution was set to 320×320 pixel.

We initialize the plane sweep dense matching cost volume dimensions zFar and zNear to the range $[h_g - t_o, h_g + b_o]$, with h_g denoting the ground level height interpolated from the terrain mesh and b_o being a fixed offset for maximum expected building height for the reconstructed city and t_o being a terrain height uncertainty offset to account for slanted ground surface or wrong terrain heights. We fix the values to b_0 = 40 m and t_0 = 10 m for that specific area. We choose a coarse discrete depth sampling Δd = 1 m in our experiments, defining the maximum achievable depth accuracy. ZNCC matching is performed with window radius r = 5. We set the outlier threshold to 0.85 to remove 85% of the detected outliers in the clustering based outlier removal step.

4.1. 3D City Model versus Dense Reconstruction

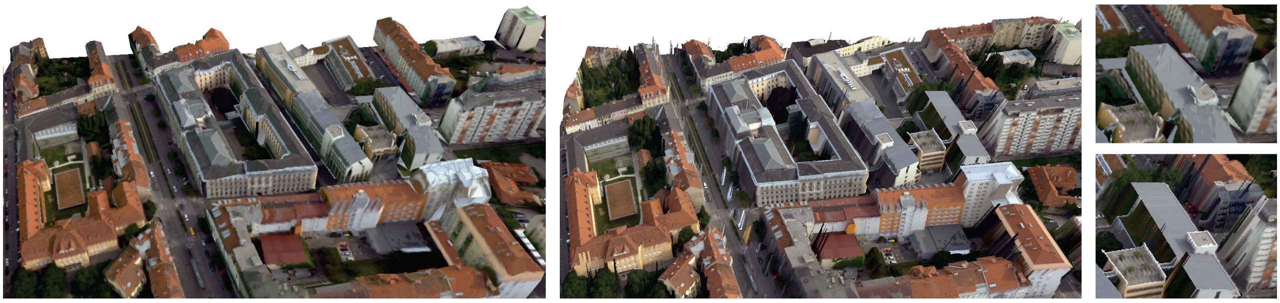
We are able to reconstruct a polygonal 3D city model with 220k triangles of an area of 1 km^2 with more than 700 buildings in less than 10 minutes on a commodity system with 3.2 GHz multi-core processor with 12GB RAM and an NVIDIA GeForce GTX 580 with 1.5GB of memory, whereas it took 6 hours to compute a dense reconstruction in full resolution based on [16, 23]. Figure 7 shows a qualitative comparison between a textured model created by our city model approximation from low resolution images and a completely dense reconstructed full resolution DSM. Although the complexity of the underlying geometry of our city model (as seen in Figure 8) is far below the degree of detail of the dense reconstruction, texturing enhances the visual quality of the model visualization to appear very similar.

The model is textured from all available views. Standard aerial images show sufficiently large viewing angles to allow the extraction of texture information for the entire model, including facades. Alternatively, oblique views such as images available in Microsoft Bing Maps Bird's eye or Google Maps and terrestrial street side images may be applied to further enhance visual quality.

Moreover, the model is free of clutter and buildings are reconstructed. Street levels are all flat, with non-building objects (e.g. cars, pedestrians, trees, etc.) removed. The 3D city model is perspectively correct, whereas the single depth map in Figure 7(b) suffers from perspective distortions. In the case of depth maps, it would be necessary to fuse several maps into one consistent DSM, which is avoided by our approach.

4.2. Accuracy Considerations and Level-of-Detail

We did a quantitative comparison for a number of hundred randomly picked sample points. Nevertheless we perform image matching at a low maximum resolution, this is sufficient to achieve reconstructed building heights with an accuracy of +/- 2 m, evaluated by comparison to a reference height model obtained by airborne laser scanning (ALS). This accuracy level is according to LOD 2 of the CityGML



(a) (b) (c)

Figure 7. A 3D city model with 220k triangles reconstructed by our algorithm in less than 10 minutes with approximated buildings from publicly available data and low-resolution aerial images is shown in (a). In contrast, (b) shows a high resolution depth map with 43M triangles which took 6 hours to compute. A comparison is given in (c): The top detailed view shows the result of our algorithm, which produces perspectively correct reconstructions with correctly textured vertical, planar facades. The bottom detailed view suffers from perspective distortions and incorrect texture due to a single view projection. Fusion of multiple depth maps into a DSM would be necessary, however this is not straight-forward.

standard, showing building models with planar facades and roof structure in Figure 8. Coarse building model geometry at this level is, when textured, already suitable for visualizations at large and medium scale. 3D city models that show buildings with flat facades only are also useful to authorities for city planning, simulations on radio signal propagation or noise pollution and natural disaster management.

Possible sources of errors include the accuracy of the SRTM elevation data for terrain modeling, since it is calculated from a filtered DSM originally containing buildings and vegetation. The absolute vertical error is reported to be up to +/- 7 m for Germany [14].

Also the credibility of voluntarily collected OSM data states a problem. The horizontal accuracy is limited due to the data capturing method using consumer grade GPS receivers with an absolute accuracy of about 10 m and the manual digitalization of building footprints from non-true-orthophotos. Moreover, the availability and completeness of OSM data in rural areas may reduce the applicability of our approach, however this is less a problem in urban environments.

Using our reconstruction approach, it would be feasible to apply corrections to the elevation and map data. Correcting OSM building footprints or giving back and adding value to OSM by updating building height tags is therefore possible.

5. Conclusion

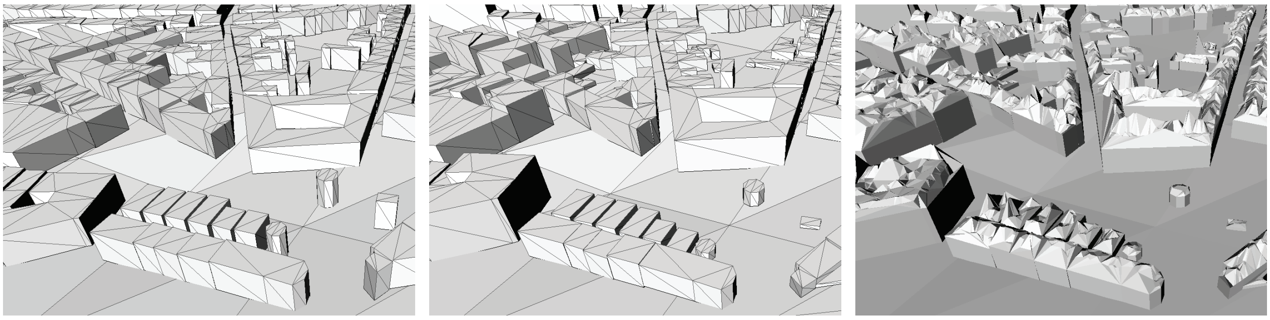
We have presented an approach for 3D city modeling comprising publicly available geographic data sources and aerial images. We showed that we can create initial three dimensional, textured surface meshes of urban scenes by using height information, building outlines from OpenStreetMap 2D map data and low-resolution aerial images for building reconstruction. Generated city models contain polyhedral building models with planar facades and approximated roof structure, satisfying CityGML Level-of-Detail 2 specification and are suited for city planning, visualization or simulations.

Our method is able to construct digital surface models without the need for fusion of multiple range scans or depth maps. A key advantage of our method is its object centered approach, creating a coarse base mesh for future refinement, if enhanced details are required. This allows algorithms to take advantage of already known geometry for applications such as initialization of dense reconstruction algorithms, visibility checks and occlusion handling, 3D model alignment and georeferencing, flight planning and obstacle avoidance for unmanned aerial vehicles (UAVs) [22] or 3D car navigation.

OpenStreetMap online data is not only a valuable source for obtaining geospatial geometric data, but provides a great resource of user-annotated content, giving meaning to objects. This allows for label transfer for classification from the web to images or 3D objects, leading to semantically interpreted maps [24], 3D models and even more holistic *virtual habitats*.

Acknowledgements

This work was supported by the Austrian Research Promotion Agency (FFG) FIT-IT project HOLISTIC (830044).



(a) (b) (c)

Figure 8. Level-of-Detail. Buildings with planar facades and flat roofs with (a) constant and (b) estimated heights assigned (LOD 1). (c) Buildings showing roof type and orientation (LOD 2).

References

[1] CGAL, Computational Geometry Algorithms Library. http://www.cgal.org. 3, 5

[2] C. Brenner and N. Haala. Rapid production of virtual reality city models. *GIS - Geo-Informations-Systeme*, 12(3):22–28, 1999. 1

[3] R. T. Collins. A space-sweep approach to true multi-image matching. *CVPR*, 1996. 4

[4] T. G. Farr et al. The Shuttle Radar Topography Mission. Technical report, Rev. Geophys., 45, RG2004, 2007. 2

[5] A. J. Flanagin and M. J. Metzger. The credibility of volunteered geographic information. *GeoJournal*, 72(3-4):137–148, 2008. 2

[6] M. Goodchild. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4):211–221, 2007. 2

[7] N. Haala. Comeback of digital image matching. *Fritsch, D. (Ed.), Photogrammetric Week 2009. Wichmann Verlag*, 2009. 3

[8] N. Haala and M. Kada. An update on automatic 3D building reconstruction. *ISPRS*, 2010. 2

[9] A. Habib, R. Zhal, and C. Kim. Generation of complex polyhedral building models by integrating stereo-aerial imagery and lidar data. *Photogrammetric Engineering and Remote Sensing*, 2010. 2

[10] A. Irschara, C. Zach, and H. Bischof. Towards wiki-based dense city modeling. In *ICCV*, 2007. 3

[11] T. H. Kolbe, C. Nagel, and A. Stadler. CityGML – OGC Standard for Photogrammetry? *Processing*, pages 265–277, 2009. 1

[12] 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*, 2010. 3

[13] F. Leberl, S. Kluckner, and H. Bischof. Collection, processing and augmentation of VR Cities. In *Fritsch, D. (Ed.), Photogrammetric Week 2009. Wichmann Verlag, Heidelberg*, 2009. 1

[14] M. Over, A. Schilling, S. Neubauer, and A. Zipf. Generating web-based 3D city models from openstreetmap: The current situation in Germany. *Computers Environment and Urban Systems*, 34(6):496–507, 2010. 2, 7

[15] T. K. Peucker, R. J. Fowler, J. J. Little, and D. M. Mark. The triangulated irregular network. *Amer Soc Photogrammetry Proc Digital Terrain Models Symposium*, 516:96–103, 1978. 3

[16] T. Pock, T. Schoenemann, G. Graber, H. Bischof, and D. Cremers. A convex formulation of continuous multi-label problems. In *ECCV*, pages 792–805, 2008. 6

[17] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz. Fast cost-volume filtering for visual correspondence and beyond. *CVPR*, 2011. 5

[18] M. Rumpler, A. Irschara, and H. Bischof. Multi-View Stereo: Redundancy Benefits for 3D Reconstruction. In *35th Workshop of the Austrian Association for Pattern Recognition*, 2011. 4

[19] A. Sampath and J. Shan. Segmentation and reconstruction of polyhedral building roofs from aerial lidar point clouds. *IEEE Transactions on Geoscience and Remote Sensing*, 48(3):1554–1567, 2010. 1

[20] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47:7–42, 2002. 5

[21] R. Wahl, R. Schnabel, and R. Klein. From detailed digital surface models to city models using constrained simplification. *Photogrammetrie, Fernerkundung, Geoinformation*, (3):207–215, 2008. 1

[22] A. Wendel, M. Maurer, A. Irschara, and H. Bischof. 3D Vision Applications for MAVs: Localization and Reconstruction. In *3DPVT*, 2011. 7

[23] C. Zach, M. Sormann, and K. Karner. High-performance multi-view reconstruction. In *3DPVT*, 2006. 6

[24] L. Zebedin, A. Klaus, B. Gruber-Geymayer, and K. Karner. Towards 3D map generation from digital aerial images. *ISPRS*, 60(6):413–427, 2006. 7