Large scale semantic 3D modeling of the urban landscape
Esteban Lopez, I.

Citation for published version (APA):
Esteban Lopez, I. (2012). Large scale semantic 3D modeling of the urban landscape

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Semantic Modeling of Tightly Packed Cities

We have discussed in the previous chapters how 3D modeling has advanced to the point of obtaining 3 dimensional point clouds. We can obtain more visually appealing models by fitting textured primitives or by obtaining dense reconstructions. This information, however accurate, still lacks an essential element: semantics.

Semantics allows you to embed meaningful knowledge into a 3D model. In particular, we are interested in modeling urban areas where we want to represent buildings. For this type of representation, we want to obtain four semantic elements: the direction of gravity, the topographical map of the city (the surface over which the city is built), the shape and height of the buildings and the geometry of the rooftops. These four elements will allow us to obtain an economical and accurate representation of the city with a particular emphasis on the buildings.

Semantics is an interesting tool that can be employed for navigation, city surveying or statistical analysis. A large scale semantic 3D model allows a robot to navigate to a specific semantic point, for instance in front of this building, parallel to that street or facing in the uphill direction, as opposed to the more traditional navigate here command. Regarding city surveying, semantics allows for planning the constructions of new buildings based on the topology underlying the city, optimal trajectories for pipes or wires based on the size, height or roof topology of buildings, simulation of disaster scenarios like flooding, analysis of sun shadows for solar energy machinery or wind flow simulations. Additionally, obtaining semantic details opens a door to obtaining statistical data on
the number of stories, roof shapes, size of facades, volume of buildings, surfaces, etc. In this chapter we present novel methods to extract semantics from a city size point cloud: direction of gravity, city surface, building shape and height and roof geometry. We employ several data sources to obtain geographical and physical information, and to obtain an accurate watertight \(^1\) model of the buildings. This information is then used to compute statistics.

The semantic modeling process is approached in four steps where the modeled city is gradually endowed with a higher degree of meaning. Firstly, the direction of gravity, also called *up vector* or vertical direction, is estimated robustly from the point cloud. Secondly, the topology of the city is estimated. This technique yields a contour map or isosurface of the ground surface of the city. Thirdly, we estimate the height of buildings’ facades. Finally, the rooftops are geometrically modeled. We integrate these four novel methods into the reconstruction pipeline described in Chapter 4.

The research presented in this Chapter was performed when the author was a visiting student at EPFL under the supervision of Christoph Strecha and Pascal Fua. We express our outmost gratitude for their support, their hospitality and fantastic views of the lake and the mountains from Lausanne.

### 6.1 Related Work

Most of the early work on semantic modeling focuses on extracting a set of high dimensional primitives to model buildings. We consider it semantic modeling because it does not simply fit geometric shapes to a set of points, but rather aims at obtaining meaningful connections between parts of the buildings.

City centers in Europe represent a modeling challenge relative to suburban areas, specially American, where buildings are usually standalone entities. European cities typically consist of buildings that were constructed at different times and of different styles, packed together to form large clusters. These clusters make single building identification and modeling very complex. Buildings are no longer single structures. Differences in height, elevation or roof topology are also present. We called these type of cities *tightly packed* cities. We aim at modeling buildings in such scenarios. In particular we want to

\(^1\) A watertight model consist of a set of polygons that intersect in such a way that they separate completely the inside and the outside of the model. They are essentially closed polyhedra.
6.1 Related Work

model the basic structure of individual buildings from ground to roof. User interaction will be permitted to narrow down the search space in data registration.

Brenner et. al. \cite{57} proposed in 2001 a mixed method that employed information from ground plans and images to obtain rough building descriptions. They approached the problem in two stages. First they divided building plans which consisted of polygonal shapes into smaller primitives. Each primitive was then fitted according to an aerial image or laser data using segmentation over depth maps. Then all primitives were assembled together to form a single building shape. Their approach was straightforward and provided rough building size reconstructions that could be manually textured. Their method produced a 2.5 dimensional reconstruction since all buildings were assumed to be resting on a single plane (they assumed a flat world).

One year later, in 2002 the work of Scholze et. al. \cite{58} focused on modeling polyhedral rooftops. They inferred the roof topology based on 3 dimensional segments obtained from images. Then they obtained a set of planes that supported those segments. Their method produced a set of consistent planes. A complete topology of the roof of a building was then obtained using five semantic labels that defined how the planes were linked together. Their method was limited by the number of labels and could not cope with arbitrarily shaped rooftops.

In 2006 Verma et. al. \cite{59} proposed a method to obtain watertight models for complete buildings from aerial laser data. They employed a decomposition technique where they fitted a set of simpler primitives to construct a complete building that fitted the point cloud as closely as possible. They then connected those simpler primitives using graphs of roof topologies. Additionally, they provided the means to obtain an estimation of the ground surface underlying the buildings. They characterized points that had a neighborhood with small change in elevation as surface points. Having identified and smoothed those points a triangular mesh was fitted to them. Their method coped well with clearly identified buildings where separation into primitives is simple, though it was not clear how the topology could be identified for more complex buildings with no rectangular shapes.

Engels et. al. \cite{60} proposed in 2007 a method for obtaining roof topologies based on laser data. They employed RANSAC to fit planes to the point cloud. Once planes were fitted to the points, plane regions were obtained using morphological filtering. They assumed that the outline of the building was given as a closed polygon that represented the
6. SEMANTIC MODELING OF TIGHTLY PACKED CITIES

facades of the building. In order to refine the roof topology they employed a refinement procedure similar to bundle adjustment. Their method relied heavily in manually set thresholds which made it unreliable for arbitrary datasets.

In 2008 Lafarge et. al. [61] presented an approach to construct a city model using building blocks. They extracted 2D-supports from the satellite digital elevation models either manually or interactively. This produced a set of 2 dimensional polygons that were then used to fit a 3 dimensional building block. The building blocks were stored in a library that aimed at covering the full set of possibilities. These blocks were combined to complete the 2D-support using Bayesian decision. They showed results on city scale reconstructions using different resolution elevation models. For lower quality models, the 2 dimensional supports were obtained manually, which defined the structure of the building in terms of block division. Roof structure was not specifically modeled and left to the selection of blocks from the library.

Elberink et. al. [62] presented in 2009 a graph theory based approach to reconstructing the geometry of buildings based on incomplete laser data. They also employed 2 dimensional polygons that denoted the outer limits of the buildings. They assumed, as previous authors did, that roof topologies could be described as a set of faces and the relations between neighboring faces. They defined some typical roof structures and stored them in a database, then a matching algorithm was applied over the data. They showed results on both complete and incomplete point clouds obtained with laser data. The buildings they reconstruct were mostly stand alone buildings with simple roof geometry and a simple polygonal outline.

More recently, Chen et. al. [63] proposed a method for modeling accurately the point cloud of a building with a set of primitives: planes, spheres and cylinders. The shapes were represented in an algebraic form and then a Least Squares method was used to fit them to the set of points. Their method was noniterative as opposed to previous work. Geometric constraints about the relation of the shapes were also introduced in the Least Squares solution to improve the reconstruction accuracy. First they obtained a normal for each point in the point cloud. Then the set of points was clustered based on their normal. Each cluster was approximated by its primitive shape selected by the distribution of normals. Then a Least Squares method was applied to fit the shapes together. They presented results using a LiDAR generated point cloud and they assumed a filtering method could be applied to extract the set of points that belonged to one
single building. As with previous methods, the shape based fitting could only cope with simple roof and building structures.

6.1.1 Discussion and Approach

Most of the methods discussed here can only cope with single building reconstructions. They employ data obtained with highly accurate laser scanners and assume that a polygonal outline is available to isolate the set of points that belong to a building. The density and resolution of the data used varies depending on the size of the environment. Some of the methods rely on libraries that contain pre-calculated models while others employ a limited set of shapes. All those primitives are then glued together to model the complete building. Most of the methods can only cope with simple roof topologies and structures. All these methods perform well over the datasets they employ. However, they do not succeed in modeling tightly packed cities such as the ones described in our urban scenarios. In such cities, the separation between buildings is far from clear and the roof topologies cannot be easily classified using the relation between faces. The methods that automatically extract the polygons that define the building outline will also fail in such scenarios due to the complexity of their structure, the differences in height, elevation and roof topology. Furthermore, save for one approach, the direction of gravity is assumed to be known and the ground surface of the city is roughly estimated or considered flat.

Most of the city centers in Europe do not fit the assumptions made in the reviewed work but rather belong to the tightly packed city style, where the modeling is more challenging. We approach the problem by employing three sources of information. We use point clouds obtained using ground and aerial images. These point clouds are less dense than the laser data and less accurate. We also employ 2 dimensional GIS data obtained by direct user input. Finally, we employ satellite images from Google. Employing these data sources we obtain a semantic model of tightly packed city centers. The steps for semantic modeling are shown in figure 6.1. The sources of information used at every step are also indicated.

First, we present a novel method to accurately estimate the direction of gravity, which allows us to obtain a vertical projection of the point clouds. Inspired by Chen et al [63] and Engels et. al [60], we employ a RANSAC based plane fitting method in a few selected points in facades to estimate the direction of gravity. With this, the
three sources of information can be registered. The direction of gravity and the point clouds are then used to obtain an accurate topological map of the underlying ground surface of the city, where in a similar way as Verma et. al. [59] we smooth the total surface by assuming a locally flat surface. Having obtained those, the height of building facades is estimated with a robust method based on the deviation of points from the facades provided by the GIS data. The remaining points belonging to the roofs are modeled through a watertight model composed of planar primitives. Instead of the plane fitting algorithm proposed by Engels et. al. [60], we employ image analysis to find the connecting points of the roof planes and a 2D mesh to obtain the connected planes similar to Scholze et. al. [58] only without limitations in the connecting possibilities. These four novelties allows us to construct semantic models of tightly packed urban areas without making any assumptions about the direction of gravity, the underlying surface or the primitives that compose each building. Finally the model is textured using the aerial image.

![Diagram of pipelines for semantic modeling of tightly packed cities using GIS, point clouds and aerial images.](image)

**Figure 6.1:** Pipelines for Semantic Modeling of tightly packed cities using GIS, point clouds and aerial images.
6.2 Sources of Information

For the task of semantic modeling we employ three different data sources: GIS data, ground based and aerial point clouds, and satellite images. In this section we discuss the extent, size and details of these sources of information.

6.2.1 GIS and the Outline Polygons

One of the fundamental steps in semantic modeling is separating the set of points that belong to buildings from the rest of the data in the point cloud. This, for certain types of buildings, can be done without assistance from further data by simply establishing certain constraints on the variation of height in the neighborhood of each point as suggested by Verma et. al. [59]. While this approach works well for standalone buildings, it is unrealistic for tightly packed city centers. Some authors [61] approach the problem by analyzing aerial depth maps and establishing boundaries with the surrounding environment with the possibility to refine them manually. This approach is equally unlikely to succeed in tightly packed urban centers.

A common approach is to assume that the outline of the building is provided. This data can be manually introduced by selecting the corners of each building in an aerial image. For large scale areas this is a daunting job. In order to isolate the set of points that most likely belong to a single building we employ GIS information, freely provided by OpenStreetMap. This data contains a set of closed polygons that represent buildings in the area (see Figure 6.2). The corners of the polygons are represented in a 2 dimensional coordinate system.

OpenStreetMap is created by users through a process of selecting corner points in satellite images. This user centered manual procedure sometimes yields outline polygons that are not accurate enough. Especially for single buildings that are composed of multiple structures[1] OpenStreetMap data tends to be a bit simplistic. For these rare occasions, we extend the GIS data by manually introducing additional points (see Figure 6.3). The GIS data provided by OpenStreetMap also contains labels that describe the use of each building. We apply user interaction only on cathedrals and palaces.

[1]For instance, a cathedral is a single building that is composed of several structures of very different heights and with very different roof topologies.
6. SEMANTIC MODELING OF TIGHTLY PACKED CITIES

6.2.2 Large Scale Point Clouds from Ground and Air Data

Using sets of point clouds obtained with lasers or LIDAR is very common. These devices provide large and dense sets of points (density depends on the distance to the objects, though typically a few hundred points per $m^2$ can be obtained), however they are expensive and difficult to operate.

Dense reconstruction approaches such as the ones discussed in Chapter 3 provide an alternative method for obtaining large sets of 3 dimensional points. Strecha. et. al. [30] propose a method to register large point clouds reconstructed using images at the ground and air level. For the data recorded at ground level, a camera is mounted in a car equipped with a GPS unit. For the aerial images a small unmanned airplane is used equipped with GPS and INS sensors. The point clouds are produced with a
6.2 Sources of Information

Figure 6.3: Palais de Rumine. BLUE: OpenStreetMap data. RED: Manually enhanced polygons.

reconstruction pipeline as described by [30] [42] and they present results that span the city center of Lausanne. We employ this sets of aerial and ground based points. The point cloud obtained from aerial images contains 6.4 million points and covers an area of 0.4 km² (an average density of 40 points per m²). The point cloud obtained from ground images contains 1 million points and covers a slightly smaller area. The points for the ground based point cloud are mostly located in the facades of the buildings where the density varies a lot throughout the cloud.

Figure 6.4: Point cloud obtained from aerial images.
6. SEMANTIC MODELING OF TIGHTLY PACKED CITIES

Figure 6.5: Point cloud obtained from ground images.
6.2 Sources of Information

6.2.3 Satellite Images

We employed satellite images obtained from Google Maps for ground and roof texturing. For the area of Lausanne we composed manually an aerial image (see Figure 6.6) of 16 megapixels and a resolution of 25 pixels per $m^2$. This composition was created through a process of tiling high-resolution images of the center of Lausanne. The image 6.6 shows the complexity of the layout and construction of the city and the intricacy of the relationships between individual buildings.

![Aerial Image obtained from Google Maps.](image)

**Figure 6.6:** Aerial Image obtained from Google Maps.
6. SEMANTIC MODELING OF TIGHTLY PACKED CITIES

6.3 Estimation of the Direction of Gravity

Estimating the direction of gravity is an essential step in creating a semantic model from a 3D reconstruction. The gravity or up vector provides information regarding the direction in which buildings are constructed and serves as an orthogonal reference for facades. For a city size scale consisting of a few hundred buildings and covering an area of $1km^2$, the theoretical direction of gravity varies $0.0025\%$ or $0.009$ degrees, within that area (assuming a perfectly round earth with all the mass located at the center). Estimating the up vector also serves to provide projections onto the ground plane. Because of this small variation, we will consider a single ground plane for the whole city. This projection of the 3D model or point cloud is used to register different datasets such as aerial, ground based or satellite images.

In order to find the direction of gravity, we assume that on average, facades are built orthogonal to the direction of gravity. Using the technique described in Section 4.6.2, we employ RANSAC to find planar patches in the point cloud obtained from ground images. Given that the point of view is at the ground level (the bottom of the facades), it is reasonable to assume that most of the reconstructed points belong to facades (see Figure 6.5). This procedure yields a set of planar polygons and their normals. We then find gravity as the vector $z^*$ that minimizes the sum of the dot products with each facade normal $n_i$:

$$z^* = \arg \max_z \sum_{i=0}^{N} |z \cdot n_i|,$$

(6.1)

with $||z|| = 1$ and $||n_i|| = 1$.

Both $z$ and the normals are normalized to length 1. Possible outliers are removed using RANSAC. In other words we find the direction that is most orthogonal to most facade normals. The Levenberg-Marquadt algorithm is used finally for refinement.

Alternative methods compute the normal of each point in the cloud based on the surrounding points and then use those normals to compute the up vector. Our method employs fewer points and outliers are removed. This provides a lower computational burden.
6.3.1 Results

For the point cloud of Lausanne containing 1 million points, 2000 planar patches are found that contain at least 200 points. Outliers are rejected using RANSAC with a fitting threshold of 0.1 meters. The nonlinear optimization converges after 3 iterations. For evaluation purposes we compare the obtained gravity vector with the up vector provided by the INS sensor of the unmanned airplane used to gather the data, obtaining an error of 0.0031 degrees, only 3 times bigger than the theoretical error based on the size of the area. Figure 6.7 shows the average error over 10 runs in the estimation of gravity as a function of the number of planar patches.

![Figure 6.7: Average error over 10 runs in the estimation of gravity as a function of the number of planar patches used.](image)

6.4 User Driven Dataset Registration

All three sources of information are in different scales and coordinate systems. Before we can model the buildings, we need to register them. We employ the direction of gravity to project the point clouds vertically obtaining 2 dimensional sets of points.
6. SEMANTIC MODELING OF TIGHTLY PACKED CITIES

GIS data and satellite images are 2 dimensional, while the point cloud is 3 dimensional. We tackle the registration problem in two stages. Firstly we use simple user input consisting of corner matches between the 2D GIS data/aerial image and a vertical projection onto a plane of the 3 dimensional point cloud. All 3D points are projected onto a plane orthogonal to the direction of gravity, which we estimate using the method described in Section 6.3 and we call the ground plane. A simple user interface displays both the GIS data with the building outlines, the aerial image and the 2D projection of the point cloud. The user then selects at least 2 pairs of corresponding building corners. Those points are used to compute the rigid 2D transformation and scaling to bring the GIS data/aerial image into the same coordinate system as the point cloud such that:

\[ \mathbf{x} = R(s\hat{x} + t), \]

where \( R \) is the 2 dimensional rotation, \( s \) is the scaling factor, \( t \) the translation and \( \mathbf{x} \) and \( \hat{x} \) the selected matching points in projection and GIS data/aerial image respectively. Rotation, translation and scaling account for a total of 4 parameters. Only two points are required to obtain the transformation in closed form. For robustness we allow the user to select more points. For every pair of points we obtain two linear equations. Compiling those equations together we obtain the transformation as a Least Squares solution using the DLT algorithm.

6.4.1 Registration Optimization

The second stage of the registration of GIS data aims at refining the Least Squares estimate. We are employing the estimation of the direction of gravity (see Section 6.3), which we assume is as accurate as possible. Therefore we fix the rotation angle around the direction of gravity. This leaves three parameters to optimize: a rotation around the direction of gravity and a 2 dimensional translation in the ground plane. Optimization in the vertical direction is not required since the GIS data is essentially 2 dimensional and we will later estimate the ground surface over which the buildings are constructed. We have two fundamentally different point clouds. The set of points generated from aerial images contains mostly points of the roofs, while the set of points generated from ground images contains mostly points of facades. We know that points that belong to a facade should project vertically near the outline of the building as described in
6.4 User Driven Dataset Registration

The GIS polygons. Points that belong to the roof should project vertically inside a slightly stretched outline polygon to accommodate for roof protrusions. Based on this, we construct an optimization procedure. The goal is to minimize:

- the accumulated distance between the projected facade points \( x_f \) and the closest outline segment \( l_c \)

- the accumulated distance between the vertices \( O_{xr} \) of the convex hull of the roof points and the closest outline segment.

Therefore, we construct a cost function \( f(x_f, x_r, L) \) such that:

\[
f(x_f, x_r, L) = w \sum_{i=0}^{N} d(x_{fi}, l_c) + (1 - w) \sum_{j=0}^{N} d(O_{x_{rj}}, l_c)
\]

where \( x_f \) is the vertical projection of the 3D points from the ground based point cloud, \( x_r \) is the vertical projection of the points from the air based point cloud, \( L \) is the polygon for a given building consisting of segments \( l \), with \( l_c \) being the closest to a given projection and \( O_{x_c} \) is the outline of the vertical projection of the rooftop points, computed as the convex hull. The weight factor \( w \) is chosen as the ratio between the number of vertices of the convex hull and the number of facade points plus vertices.

This method aims at fitting the GIS data to the vertical projection of the city such that the outlines of the buildings fit best with the actual point cloud. It is based on the assumption that the number of points in the cloud that belong to the ground is low with respect to the number of points that belong to facades or rooftops.

The distance between one projected point \( x_{fi} \) and the closest segment \( l_c \) is defined as the shortest distance between a point and line segment.

The distance between two polygons is defined as the sum of distances between each of the \( M \) segments of \( O_{xr} \), and the closest segment in \( L \), defined as \( l_{cO} \). The closest segment is defined as the one with minimal distance between the centroids of the two segments. This measure is motivated by the fact that the buildings are roughly aligned by user input and the points that belong to the roof are expected to fall within the area enclosed by the GIS polygon, therefore most of the segments from the convex hull will be very close to the segments of the GIS polygon (see figure 6.8).
6. SEMANTIC MODELING OF TIGHTLY PACKED CITIES

Figure 6.8: LEFT: facade points project near the building outline. RIGHT: roof points project inside the building outline.

The error function is represented in Figure 6.9. Whiter areas represent a lower error value. The value for each point (either point of the facade or a vertex of the convex hull) is obtained by projecting the point in the map. The further away from the GIS polygon, the larger the error. The optimization is performed for all buildings at once.

Figure 6.9: Representation of the error function for optimization.
6.5 Ground Surface Reconstruction

Once the direction of gravity is obtained, we estimate the topological map of the city. For this we employ a surface fitting approach that works from the bottom up. All points from the point clouds must lie on top or above the surface of the city. We start the procedure with the ground plane. This plane is sampled using a square grid with a grid constant of 5 meters and located at the height of the lowest point in the cloud. For reference, we call the sampled plane or grid the heightmap since it maps the height of the ground surface of the city model. This map can be displayed as a 2D image similar to a contour map.

Every node of the grid is set to a height equivalent to the lowest 3D point (from the point cloud) contained in a cell of 5 meters along each side centered at the node. This produces an irregular fitting where nodes are initialized only where evidential data is found. Additionally, we build a grid of the same size as the heightmap that we use to store weights that support the evidence for the estimation of height. We call this the weightmap and it is used later on the height propagation procedure. The weight is initialized as the ratio between the number of points contained in the 3D cell and the total number of points in the cloud.

After the height-map has been initialized, we smooth out the sudden changes in height. These sudden changes in height are caused by lack of evidence such as empty areas in the point cloud. We define the maximum declination as the absolute difference between one cell with evidence and the minimum height of the eight surrounding cells. If the maximum declination is above a certain threshold, then the height of the cell is set of the minimum of the neighboring cells. This threshold represents the maximum ground elevation change in the area. Additionally, small isolated structures are removed (see Figure 6.10 - BOTTOM-LEFT) since they most likely represent small reconstructions that do not account for a full building or even a large part of one, for instance when barely any points are reconstructed or when points belong to a tree or a small statue. This is achieved by finding small patches in the weight-map and setting them to zero.

The final step consists of smoothing the heightmap based on the weightmap. We apply a convolution filter constructed with the weights of the surrounding cells.
6. SEMANTIC MODELING OF TIGHTLY PACKED CITIES

Figure 6.10: Ground surface estimation process: TOP-LEFT. Height-map initialization. TOP-RIGHT. Large declination smoothing. BOTTOM-LEFT. Isolated patch removal. BOTTOM-RIGHT. Final smoothed ground height.

6.5.1 Results

We apply our proposed method to the point cloud (both air and ground) of the city of Lausanne. The complete set of 7.4 million points is reduced using density driven sampling to obtain 0.5 million points. Figure 6.10 shows the four stages of the estimation process where buildings are gradually smoothed out of the resulting topological map. Figure 6.11 compares the obtained topological map with an isoline map provided by the council of Lausanne. Red lines represent the same height. The hill where the cathedral is located is accurately recovered along with the boundaries with Rue Saint-Martin.
Figure 6.11: Validation of the estimated ground surface.
6. SEMANTIC MODELING OF TIGHTLY PACKED CITIES

6.6 Estimation of the Height of Buildings

Once the surface of the city is estimated, we obtain the height of the facade by observing the change in the average distance between the sparse points and the facades. Since the distance from point to facade is the orthogonal segment that spans from the point to the facade and we have estimated the direction of gravity, this is equivalent to computing the distance in the projected space. One building commonly consists of a multiple set of facades so we only consider the distance of each point to the closest facade.

We propose a method that works in three consecutive stages. First we construct a 2D vector \( V = [z_i, d_i] \) containing the height \( z_i \) of each sparse point \( i \) and the distance \( d_i \) to the closest facade normalized by the distance to the centroid of the 2D outline polygon. Over this vector a 2 dimensional histogram is constructed (see Figure 6.12 TOP-LEFT). Then we multiply this distribution by an exponential function (see Figure 6.12 TOP-RIGHT), obtaining a distribution where relevance of points is graded as they are further away from the facades (see Figure 6.12 BOTTOM).

We then sum up over the distance and obtain a 1 dimensional distribution (see Figure 6.13). Finally, we compute the derivative and obtain the roof line as the minimum height at which the derivative is higher than 20% of the maximum value. This value represents the highest point in the facade and the point where the roof begins. We employ the 20% threshold to represent the noise present in the point cloud. The idea is to identify the first instance where points start to deviate from the facade due to architecture rather than noise.

6.6.1 Results

The facade estimation algorithm is applied over all buildings present in the GIS data for which sufficient evidence exists to support the separation between roof and facade. The GIS data of Lausanne contains 937 buildings. For 742 of them sufficient evidence is found to obtain an estimate for the facade height in the point cloud. Three different building models are shown in figure 6.14 along with the corresponding roof points. In all three the height of the facade is accurately estimated allowing for a clear cut separation between roof points and facade points. Since the building facade model is entirely based on the accuracy of the GIS data, for buildings where the point cloud does not correspond with the GIS building outline, the facade cannot be correctly estimated.
6.6 Estimation of the Height of Buildings

Figure 6.12: Three stage procedure for obtaining the height of facades. TOP-LEFT: distribution of point-facade distances over height and distance. TOP-RIGHT: exponential function used to grade points based on their distance to the facades. BOTTOM: resulting graded distribution.

Figure 6.15 shows the estimated height of a building together with an image of the real building.
6. SEMANTIC MODELING OF TIGHTLY PACKED CITIES

Figure 6.13: Summed up distribution of the height of a single building (BLUE) and derivative (RED).

Figure 6.14: Estimation of the height of facades for 3 different buildings.
Figure 6.15: Comparison between estimated facade height and real building.
6.7 Rooftop Modeling

Having obtained an estimation of the height of the facades, we propose in this section an automatic method for estimating a geometric model of the rooftop. We consider the roof as the outline polygon that coincides with the outline of the building, and a set of height control points that define the 3 dimensional shape. Those control points must lie inside the outline polygon.

Some authors consider the rooftop as a collection of planes and a set of predefined configurations they can be found in. This technique, together with the pre-calculated models approach, cannot be applied to highly irregular building outlines such as the ones found in city centers.

We approach the geometric modeling problem in four stages. First, the control points of the roof are introduced. Second, the height of those points is estimated. Third, the set of planes that form the geometric roof is determined. Finally, a margin approach is employed to reduce the number of faces of the geometric model.

6.7.1 Finding Control Points

The first step consist of introducing control points to be used in the geometric structure. These points relate to corners, ridges and similar key points in the structure of the roof. In order to automatically estimate those points, we employ the 17Mpx aerial image. We compute the corners using the Harris method for the part of the image that lays inside the building outline polygon. The motivation to use a corner detector is the fact that key points in the roof structure are very likely to represent places where two or more planes intersect. This confluence of planes implies a significant difference in the direction of the normal, which in turn will produce a difference in lighting. This difference will be further enhanced by changes in shadows and materials. This method is very likely to recover most of the relevant points in the roof structure, though it will also recover nonrelevant ones such as points where the intensity in aerial image varies due to material changes.

6.7.2 Estimating the Height of Control Points

Once the control points are determined, their height is estimated using the points of the aerial point cloud that lie inside a polygon that slightly bigger than the outline of
6.7 Rooftop Modeling

Figure 6.16: LEFT: The aerial image and the outline of the polygon of a building is shown. Red crosses represent the Harris corners found in the image. Blue circles are the corners that lie inside the polygon. RIGHT: triangulated vertices.

the building. The polygon is bigger to accommodate for small errors in the registration. We employ a polygon 30% bigger, the more accurate the registration, the lower the number. The height of each control point is estimated using a similar technique as the one employed in the estimation of the ground surface. For every control point, a height is set equal to the lowest height of the 3D points contained just above the point in the direction of gravity. If no points are present, the height is set to the height of the building, placing the control point in the roof plane.

6.7.3 Finding Planes

We want to obtain a geometrical and watertight 3D model of the rooftop, therefore we now need to calculate the planes that intersect in the estimated control points.
6. SEMANTIC MODELING OF TIGHTLY PACKED CITIES

For this we employ the Delaunay triangulation method on the vertical projection of the points. This yields a set of triangles that cover the complete rooftop making a watertight geometric model. The vertices of the triangles are the estimated control points and the vertices of the outline polygon of the building.

6.8 Modeling the city of Lausanne, Results and Discussion

In this Chapter we have proposed methods for estimating certain semantic aspects of the reconstruction of a city. For our experiments we obtained two 3D point clouds of the city of Lausanne from Strecha et. al. The aerial point cloud consisted of 6.3 million points reconstructed from aerial images recorded with an UAV (Unmanned Aerial Vehicle). GPS and INS sensors were also used to obtain the reconstructed points. In this case the recorded INS information provided an estimation for the direction of gravity. The ground based point cloud consisted of 1 million points reconstructed using images recorded with a ground vehicle.

GIS data was also obtained from the Open Street Map project. The outline polygons for 937 buildings were obtained. Using Google Maps, a 17 Mpx image was composed of the area covered by the point clouds.

Figure 6.17 shows a top view of the three sources of information registered. Figure 6.18 shows the aerial image (see Figure 6.6) used for texturing the ground plane. The registered buildings obtained from GIS are also displayed with a fixed height. The same procedure can be applied over the computed ground surface and it is shown in Figure 6.19.

Once the point cloud has been registered and the ground surface has been estimated, the height of each building can also be estimated. Two views of the entire city center of Lausanne are shown in Figure 6.20.

6.8.1 Statistical Modeling of Building Properties

Obtaining a 3 dimensional model of buildings by estimating the height of their facades is useful for visualization. Within those models an additional layer of information is inferred, namely the height of the buildings, the area they occupy and their volume. We employ these properties to obtain 3D visualizations of the buildings. Details of the
6.9 Conclusions

Figure 6.17: Top view of the registered sources of information. Green shows the aerial point cloud. Red shows the ground based point cloud. Blue shows the GIS data.

ground surface can also be visualized, such as height or deviation of the normal from the direction of gravity.

Figures 6.21, 6.22, and 6.23 show the three building properties mentioned above. In these models, identifying key buildings is easy and has many potential uses in city surveying or planning.

6.9 Conclusions

In this Chapter we have developed the methods and techniques to obtain a semantic model of a city based on three sources of information: point clouds, GIS data and aerial images. The goal was to obtain a model with valuable information for city surveying tasks.

In a stepwise approach, we have increased the level of complexity of the simple point clouds. First, we showed a robust and accurate method to estimate the direction of gravity. This is a key element in order to put the model in a realistic reference system. Based on this direction, an estimation of the topological map of the city was obtained. For this we developed a multistep grid based approach that created a smooth surface over which buildings could be placed. Using freely available GIS data, we developed
a method to estimate the height of the facades of buildings. This, together with the estimated city surface, allowed us to place the GIS data in a truly 3 dimensional context, placing watertight buildings in a realistic environment. This produced a set of flat roof models for the buildings. We improved those by creating a geometric model of the rooftops. For this, a set of control points was calculated based on the aerial image of the building. Then the height of those control points was estimated according to the point cloud, obtaining a watertight geometric model of the rooftops. Finally, a texture can be applied to both the surface and the roof models, obtaining therefore a rich and visually appealing representation of the city that can be seen in figure 6.24.

The information obtained during the modeling process can also be used for semantic visualization. We have shown visualizations of the height, area and volume of the buildings in a complete city.

There are two challenging aspects of the modeling process. the geometric characterization of the rooftops and the evaluations of the methods. The geometric characterization is difficult due to the limited accuracy of the aerial reconstruction and the complexity of the roof structure in dense city centers such as Lausanne. More accurate aerial images and alternative control point detectors might be used to obtain better geometric descriptions. The second challenging aspect is the evaluation of the methods. Obtaining
Figure 6.19: Google image used for texturing the estimated ground surface.

ground truth for the semantic elements we obtain is difficult and costly.
Figure 6.20: Views of the model of the city center of Lausanne where the height of the buildings has been estimated.
Figure 6.21: Area: Model of the city of Lausanne, color is assigned to buildings based on the area they occupy, from smaller (blue) to bigger (red).

Figure 6.22: Height: Model of the city of Lausanne, color is assigned to buildings based on their height, from smaller (blue) to bigger (red).
Figure 6.23: Volume: Model of the city of Lausanne, color is assigned to buildings based on the volume they occupy, from smaller (blue) to bigger (red).
Figure 6.24: Two views of the complete model of the city of Lausanne. Texture has been applied to the rooftops and the ground surface.