Large scale semantic 3D modeling of the urban landscape

Esteban Lopez, I.

Citation for published version (APA):
Summary

The objective of this thesis was to research and develop the methods and techniques for obtaining large scale semantic models of urban areas. In particular we were interested in fast and accurate techniques that do not rely heavily on numerical optimization. To achieve this, we focused on employing monocular cameras to record the environment due to their simplicity, flexibility and low cost.

In Chapter 1 we presented the field of 3D reconstruction and discussed the camera paradigm. We also introduced by means of intuition the reconstruction procedure that aims at recovering the 3D structure recorded in images. The challenges ahead were also introduced, namely how to design a reconstruction pipeline, how to recover the scale between different images and how to model an urban scene semantically.

In Chapter 2 we introduced some visual geometry concepts. We described the camera imaging process by means of the pinhole camera model. We also described a more generic camera where lenses are used to improve the quality of the image. We chose a coordinate system where objects, images and cameras can be referenced. These concepts, along with basic concepts in computer vision serve as the groundwork of the methods we have developed.

In Chapter 3 we discussed the history and state-of-the-art in the reconstruction literature, concluding that the most advanced methods solve the reconstruction problem with a numerical optimization procedure. This black box approach attempts to establish constraints between camera positions and image features in order to minimize a cost function, solving then the camera positions and obtaining a 3D reconstruction. While
these methods work well, we believe that more attention should be devoted at each of
the individual steps of the reconstruction procedure.
Thus in Chapter 4 we proposed a 6-step reconstruction pipeline. We discussed each of
the steps for the task of 3D urban reconstruction with the goal of obtaining a fast an
accurate reconstruction pipeline. We treated in detail the first five steps, reasoning and
motivating our choices of algorithms. Semantic Modeling was left a later chapter. We
provided an overview of available methods for each of the steps. With respect to the
estimation of the camera motion, we described the scale challenge, which is treated in a
separate chapter. In the end, we provided a complete reconstruction pipeline for urban
scenes where no large numerical optimization is required.
In Chapter 5 we described in detail the local scale problem. We reviewed the current
solutions, which typically involve using Image-to-World methods or real life observations
regarding the speed of the camera or its position with respect to the ground. We
began with definitions regarding the error present in image features, points in space
and camera rotation and translation. We also detailed our camera position estimation
algorithm based on our discussion and algorithmic choices from Chapter 4. Then we
derived a Least Squares closed form solution for estimating the scale where only one
Image-to-World match is required. Furthermore, there are different ways to solve the
system of equations. We proposed to solve the scale $s$ using only the motion of features
in the direction of $x$ (Method 1), which is the direction in which the camera moves
the most in an urban environment. Then we performed a first order error propagation,
which allowed us to compute the error of the estimation of the local scale $s$. Finally, we
proposed a Maximum Likelihood estimation to obtain an optimal scale. We provided
a comparison with state-of-the-art methods, where we prove the superiority of our
method.
In Chapter 6 we focused on extracting semantics for an urban 3D model. In partic-
ular, we wanted to obtain the direction of gravity, the surface over which the city is
constructed, a watertight model of the buildings from ground to roof line (where the
facade meets the roof), and finally a geometric model of the roof tops. We proposed
to use three sources of information: a ground-and-air based point cloud reconstruction,
an aerial image obtained from Google Maps and a GIS map obtained from Open Street
Map. Firstly, we proposed a fast RANSAC-based method for estimating the direction
of gravity (or the vertical direction). We assumed that, on average, the facades in the
city are built perpendicular to this direction. We compared the resulting direction with the direction provided by an INS sensor, resulting in an error of 0.0031 degrees. Secondly, we proposed a four-stages method to estimate the topological map of the city. Our method produced an accurate elevation map of the ground over which the city is built, which we compared with an isoline map available online. Thirdly, we proposed a method to estimate the height of the roof line of each building. The method works by means of observing how the distribution of facade-pointcloud point distance changes with height. The resulting heights were applied over the complete set of buildings of a city. Finally, we modeled the rooftops geometrically by analyzing the aerial images of the buildings, obtaining a complete watertight model for each building.

In Chapter 7 we presented our toolbox FIT3D, which is a Matlab implementation of the work we have developed for urban 3D reconstruction. We compared the toolbox with other well known available packages and showed some examples of the functionalities contained within the toolbox.