Geometric and semantic indoor environment interpretation

Lavinia S. Runceanu

Institute for Photogrammetry (ifp), University of Stuttgart, Germany
Geschwister-Scholl-Straße 24D, D-70174 Stuttgart
Forename.Lastname@ifp.uni-stuttgart.de

A large variety of systems and applications are offering mapping, localization and navigation services for outdoor environment. However, people spend most of their time indoor, where there is a lack in digital maps and where conventional GPS services do not work. Despite this need in indoor navigation applications, the developments in the field of augmented and virtual reality, including also game industry, and Building Information Modeling (BIM), are also requiring intelligent indoor models, especially for public buildings.

In order to obtain the needed models, an innovative solution is to use data coming from different sensors. Being well suited for mapping textureless surfaces, often present in indoor environments, active systems are mostly used for satisfying the indoor mapping request. Improvements in sensor design and technology as well as in the used algorithms have contributed to the large variety of systems used for 3D data collection today. In order to increase the mapping efficiency and at the same time to reduce the mapping costs, these systems integrate laser scanners, cameras and sometimes inertial measurement units. Also, they adapt their design to the indoor space, being built as a trolley or as a backpack (e.g. NavVis M3 Trolley, Leica Pegasus). Another step further is made by the availability of devices integrating depth cameras, at a consumer-level. These platforms are using low cost sensors which made them affordable to the general public (e.g. Phab 2 Pro Phone, Google Tango Tablet, Microsoft Kinect). All this progress enables unexperienced users to contribute to the indoor mapping request, but in the same time this rises new challenges which need to be overcome.

For many public buildings there is available 3D data coming from traditional systems, like laser scanner. Even if, for the main majority of the research work, this data is considered to be a reference, it is still not perfect. Having a limited number of static stations and being necessary to align and geo-reference the multiple sessions, also this data is affected by errors and incomplete. With the availability of 3D data coming from other sources, there would be possible to geometrically improve the final 3D model in terms of accuracy and completeness by performing an optimal data fusion. This should be optimal, namely by enhancing the model with data coming from low-cost sensors only where the data coming from the laser scanner is strongly affected by registration errors or gaps, for example. This data fusion could be very useful not only for accurately reconstructing public buildings, but also for efficiently monitoring a building construction side, preventing construction errors and keeping the BIM up-to-date to possible changes.

The 3D fused data, coming from different sources it may still be noisy to some level. Also, the complexity of an indoor environment may cause problems in individually detecting the meaningful objects, aimed to be reconstructed. A solution for these issues would be to make use of a robust algorithm in the form of a grammar, capable to describe 3D structures and semantic information. An important feature of the grammar should be its capability of learning and predicting the geometric and semantic information from the available data. If the acquired data, contains more than one room, the individual rooms (the separating walls) should be detected and modelled. Also, beside the walls, inside a room it would be possible to detect

and model the windows and doors, but also other objects like chairs, tables, cupboards, etc. According to the modelled objects, also the room destination should be determined. For doing that, the acquired 3D data needs to be automatically interpreted. For the automatic 3D data interpretation, different segmentation approaches are normally used. The results of the interpretation step, improved by the use of geometric and topological conditions present in indoor environments, can be used as a knowledge base for automatically inferring the grammar rules. However, due to the heterogeneity of the possible 3D entities from an indoor environment, it would be necessary not only to differentiate between different objects, but also to identify the object type to which that region of the 3D data belongs. On the one side, different object interpretation approaches could be used, adapted to the object to be detected. On the other side, machine learning technique, like Convolutional Neural Networks (CNN), are suitable for automatically interpreting the indoor space and determining the corresponding objects within it.