

Reconstructing 3D City Models by Merging Ground-Based and Airborne Views

Christian Frueh and Avidesh Zakhori¹

Video and Image Processing Lab
University of California, Berkeley
Cory Hall, Berkeley, CA 94720
{frueh,avz}@eecs.berkeley.edu

Abstract. In this paper, we present a fast approach to automated generation of textured 3D city models with both high details at ground level, and complete coverage for bird's-eye view. A close-range facade model is acquired at the ground level by driving a vehicle equipped with laser scanners and a digital camera under normal traffic conditions on public roads; a far-range Digital Surface Map (DSM), containing complementary roof and terrain shape, is created from airborne laser scans, then triangulated, and finally texture mapped with aerial imagery. The facade models are registered with respect to the DSM by using Monte-Carlo-Localization, and then merged with the DSM by removing redundant parts and filling gaps. The developed algorithms are evaluated on a data set acquired in downtown Berkeley.

1 Introduction

Three-dimensional models of urban environments are useful in a variety of applications such as urban planning, training and simulation for disaster scenarios, and virtual heritage conservation. A standard technique for creating large-scale city models in an automated or semi-automated way is to apply stereo vision techniques on aerial or satellite imagery [5, 10]. In recent years, advances in resolution and accuracy have rendered airborne laser scanners suitable for generating Digital Surface Maps (DSM) and 3D models [1, 9, 12]. There have been several attempts to create models from ground-based view at high level of detail, in order to enable virtual exploration of city environments. Most common approaches involve enormous amounts of manual work, such as importing the geometry obtained from construction plans. There have also been attempts to acquire this close-range data in an automated fashion, either using stereo vision [3] or laser scanners [13, 14]. These approaches, however, do not scale to more than a few buildings, since data has to be acquired in a slow stop-and-go fashion.

¹ This work was sponsored by Army Research Office contract DAAD19-00-1-0352.

In previous work [6, 7], we have developed an automated method capable of rapidly acquiring 3D geometry and texture data for an entire city at the ground level. We use a vehicle equipped with fast 2D laser scanners and a digital camera while driving at normal speeds on public roads, hence acquiring data continuously rather than in a stop-and-go fashion. In [8], we have presented automated methods to process this data efficiently, in order to obtain a highly detailed model of the building facades in downtown Berkeley. These facade models however, do not provide any information about roofs or terrain shape. In this paper, we will describe merging the highly detailed facade models with a complementary airborne model obtained from a DSM, in order to provide both the necessary level of detail for walk-thrus and the completeness for fly-thrus. The outline of this paper is as follows: Section 2 describes the generation of a textured surface mesh from airborne laser scans. Section 3 summarizes our approach to ground-based model generation and model registration. We propose a method to merge the two models in Section 4, and in Section 5, we present results for a data set of downtown Berkeley.

2 Textured Surface Mesh from Airborne Laser Scans

In this section, we describe the generation of a texture mapped airborne mesh from airborne laser scans. First we generate a Digital Surface Map (DSM), i.e. a regular array of altitude values, by resampling the unstructured point cloud of airborne laser scans, and filling remaining holes in the DSM by nearest-neighbor interpolation. This DSM can be utilized for both generating a model from airborne view and localizing the ground-based data acquisition vehicle. Since the DSM contains not only the plain rooftops and terrain shape, but also many other objects such as cars, trees, ventilation ducts, antennas, and railings, the terrain and roofs look “bumpy” and edges are jittery. In order to obtain a more visually pleasing reconstruction of the roofs, we apply additional processing steps: The first step is aimed at flattening “bumpy” rooftops. To do this, we first apply to all non-ground pixels a region-growing segmentation algorithm based on depth discontinuity between adjacent pixels. Small, isolated regions are replaced with ground level altitude, in order to remove objects such as cars or trees in the DSM. Larger regions are further subdivided into planar sub-regions by means of planar segmentation. Then, small regions and sub-regions are united with larger neighbors by setting their z values to the larger region’s corresponding plane. This procedure is able to remove undesired small objects from the roofs and prevents rooftops from being separated into too many cluttered regions. The second processing step is intended to straighten jittery edges. We re-segment the DSM, detect the boundary points of each region, and use RANSAC to find line segments that approximate the regions. For the consensus computation, we also consider boundary points of surrounding regions, in order to detect even short linear sides of regions, and to align them consistently with surrounding buildings; furthermore, we reward an additional bonus consensus score if a detected line is parallel or perpendicular to the most dominant line of a region. For each region, we obtain a set of boundary line segments representing the most important edges, which are then smoothed out. For all other boundary parts, where a proper line approximation has not been found, the

original DSM is left unchanged. Since the DSM has a regular topology, it can be directly transformed into a structured mesh by connecting each vertex with its neighboring ones. Using the Qslim mesh simplification algorithm [4] to reduce the number of triangles, we obtain a mesh with about 100,000 triangles per square kilometer at its highest level of detail. Fig. 1 shows the mesh for a downtown Berkeley area with and without processing, respectively.

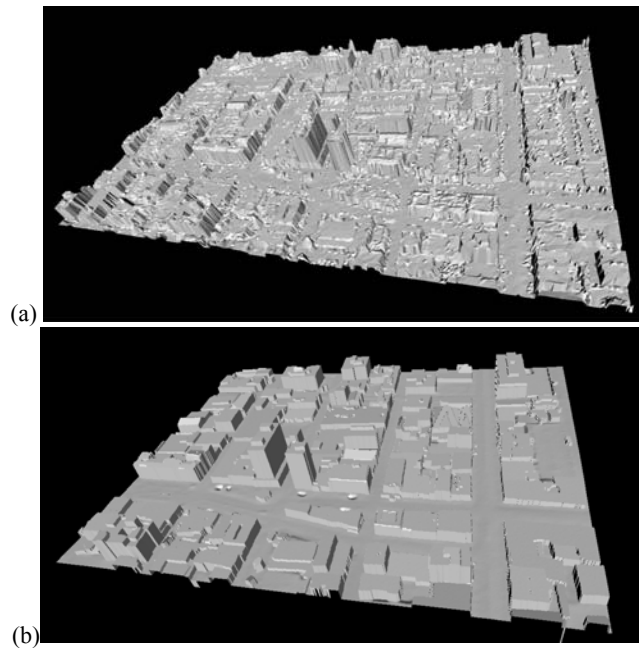


Fig. 1. Airborne surface mesh; (a) without processing, (b) after applying postprocessing steps

The reduced mesh can be texture mapped in a semi-automatic way using aerial images with unknown pose: Correspondence points are manually selected in both the aerial photo and the DSM, taking a few minutes per image, and hence about an hour for twelve high-resolution images covering the entire downtown area of Berkeley, which is about a square kilometer. Then, the image pose is automatically computed and the mesh is texture mapped: Specifically, a location in the DSM corresponds to a 3D vertex in space, and can be projected into an aerial image for a given camera pose. We utilize Lowe's algorithm to compute the optimal camera pose by minimizing the difference between selected correspondence points and computed projections [11]. At least 6 correspondence points per picture are needed to solve for the pose, but in practice about 10 points are selected for robustness. After the camera pose is determined, for each geometry triangle, we identify the corresponding texture triangle in an image by projecting the corner vertices according to the computed pose. Then, for each mesh triangle the best image for texture mapping is selected by taking into account resolution, normal vector orientation, and occlusions.

3 Ground-Based Modeling and Model Registration

In previous work, we have developed a mobile data acquisition system consisting of two Sick LMS 2D laser scanners and a digital color camera with a wide-angle lens. This system is mounted on a rack on top of a truck, enabling us to obtain measurements that are not obstructed by objects such as pedestrians and cars. Both 2D scanners face the same side of the street, one mounted horizontally, the other one vertically, and they are synchronized by hardware signals. The data acquisition is performed in a fast drive-by rather than a stop-and-go fashion, enabling short acquisition times limited only by traffic conditions. In our measurement setup, the vertical scanner is used to scan the geometry of the building facades as the vehicle moves, and hence it is crucial to determine the location of the vehicle accurately for each vertical scan. In [6] and [7], we have developed algorithms to globally register the ground-based data with an aerial photo or a DSM as a global reference. These algorithms use Monte-Carlo-Localization (MCL) in order to assign a 6 degree-of-freedom global pose to each laser scan and camera image. After the localization, we apply a framework of automated processing algorithms to remove foreground and reconstruct the facades. As described in [8], the path is segmented into easy-to-handle segments to be processed individually. The further steps include generation of a point cloud, classification of areas as facade versus foreground, removal of foreground geometry, filling facade holes, triangulation and texture mapping [8]. As a result, we obtain facade models with centimeter resolution and photo-realistic texture.

4 Model Merging

Previous approaches for fusing meshes, such as sweeping and intersecting contained volume [2], or mesh zippering [13], require a substantial overlap between the two meshes. Additionally, they work only if the two meshes have similar resolutions. Due to the different resolution and complimentary viewpoints, these algorithms cannot be applied. In our application, it is reasonable to give preference to the ground-based facades wherever available, and use the airborne mesh only for roofs and terrain shape. Rather than replacing triangles in the airborne mesh for which ground-based geometry is available, we consider the redundancy before the mesh generation step in the DSM: for all vertices of the ground-based facade models, we mark the corresponding cells in the DSM. We further identify and mark those areas, which our automated facade processing in [8] has classified as foreground. These marks control the subsequent airborne mesh generation from DSM; specifically, during the generation of the airborne mesh, (a) the z value for the foreground areas is replaced by the ground level estimate from the DTM, and (b) triangles at ground-based facade positions are not created. Fig. 2(a) shows the DSM with facade areas marked in red and foreground marked in yellow, and Fig. 2(b) shows the resulting airborne surface mesh with the corresponding facade triangles removed. The ground-based facade models to be put in place do not match the airborne mesh perfectly, due to their different resolutions and capture viewpoints. In order to make mesh transitions less noticeable, we

fill the gap with additional triangles to join the two meshes: Our approach to creating such a “blend mesh” is to extrude the buildings along an axis perpendicular to the facades, and then shift the location of the “loose end” vertices to connect to the closest airborne mesh surface. The blend triangles are finally texture-mapped with the texture from the aerial photo, and as such, they attach at one end to the ground-based model, and at the other end to the airborne model, thus reducing visible seams at model transitions.

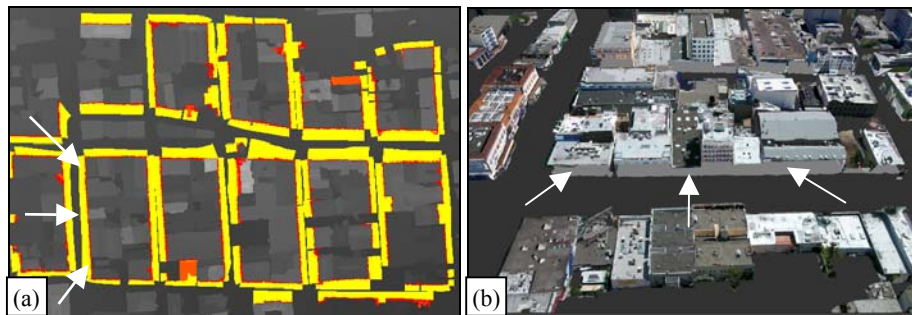


Fig. 2. Removing facades from airborne model; (a) marked areas in DSM; (b) resulting mesh with corresponding facades and foreground objects removed. The arrows indicate an example for corresponding locations.

5 Results

We have applied the proposed algorithms on a data set for downtown Berkeley. The airborne laser scans have been acquired in conjunction with Airborne 1, Inc. of Los Angeles, CA; the entire data set consists of 48 million scan points and from these scans, we create a DSM with a cell size of 1 m by 1 m square. The ground-based data has been acquired during two measurement drives in Berkeley: The first drive took 37 minutes and was 10.2 kilometers long, starting from a location near the hills, going down Telegraph Avenue, and in loops around the central downtown blocks; the second drive took 41 minutes and was 14.1 kilometers long, starting from Cory Hall and looping around the remaining downtown blocks. A total of 332,575 vertical and horizontal scans, consisting of 85 million scan points, along with 19,200 images, were captured during those two drives.

Applying our MCL approach, we register the driven path and hence the ground-based scan points globally with the DSM. Fig. 3(a) shows the driven paths superimposed on the airborne DSM, and Fig. 3(b) shows the ground based horizontal scan points for the corrected paths in a close-up view. As seen, the two paths and the horizontal scan points are geo-referenced and match the DSM closely. Note that our registration method is agnostic to the way the DSM is generated. Even though in this paper we generated our DSM from airborne laser scans, our approach would also work with DSMs created from ground plans, aerial images, or SAR data.

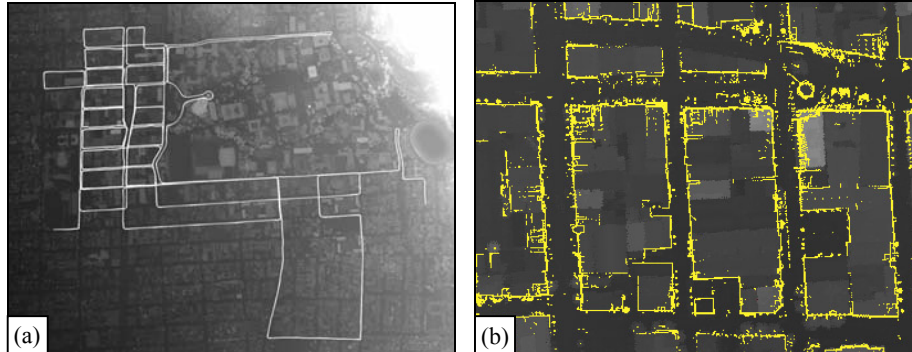


Fig. 3. Registration of the ground-based data; (a) driven paths superimposed on top of the DSM after using Monte-Carlo-Localization; (b) horizontal scan points for the registered paths

Using the processing steps described in [8], we generate a facade model for the downtown area as shown for 12 city blocks in Fig. 4. Note that the acquisition time for these blocks was only 25 minutes; this is the time that it took to drive the total of 8 kilometers around the blocks under city traffic conditions. Applying the model merging steps as described in Section 4, we combine the two modalities into one single model. Fig. 5(a) shows the resulting fused model for the looped downtown Berkeley blocks from the top, as it appears in a fly-thru, and Fig. 5(b) shows the model as viewed in a walk-thru or drive-thru. As seen, the combined model appears visually pleasing from either viewpoint.



Fig. 4. Reconstructed facade models for the downtown Berkeley area



Fig. 5. Fused 3D model of downtown Berkeley; (a) bird's eye view, (b) walk-thru view

Our proposed approach to city modeling is not only automated, but also fast from a computational viewpoint: the total time for the automated processing and model generation for the 12 downtown blocks is around 5 hours on a 2 GHz Pentium-4 PC. Since the complexity of all developed algorithms is linear in area and path length, our method is scalable and applicable to large environments.

6 Conclusion and Future Work

We have presented a method of creating a 3D city model suitable for walk- and fly-thrus by merging models from airborne and ground-based views. Future work will address (a) complete automation of aerial imagery registration, (b) adding 3D and 4D

foreground objects such as trees, traffic signs, cars and pedestrians to the ground-based model, and (c) rendering issues.

References

1. C. Brenner, N. Haala, and D. Fritsch: "Towards fully automated 3D city model generation", Workshop on Automatic Extraction of Man-Made Objects from Aerial and Space Images III, 2001
2. B. Curless and M. Levoy, "A volumetric method for building complex models from range images", SIGGRAPH, New Orleans, 1996, pp. 303-312
3. A. Dick, P. Torr, S. Ruffle, and R. Cipolla, "Combining Single View Recognition and Multiple View Stereo for Architectural Scenes", Int. Conference on Computer Vision, Vancouver, Canada, 2001, pp. 268-74
4. M. Garland and P. Heckbert, "Surface Simplification Using Quadric Error Metrics", SIGGRAPH '97, Los Angeles, 1997, pp. 209-216
5. D. Frere, J. Vandekerckhove, T. Moons, and L. Van Gool, "Automatic modeling and 3D reconstruction of urban buildings from aerial imagery", IEEE International Geoscience and Remote Sensing Symposium Proceedings, Seattle, 1998, pp. 2593-6
6. C. Frueh and A. Zakhor, "Fast 3D model generation in urban environments", IEEE Conf. on Multisensor Fusion and Integration for Intelligent Systems, Baden-Baden, Germany, 2001, pp. 165-170
7. C. Frueh and A. Zakhor, "3D model generation of cities using aerial photographs and ground level laser scans", Computer Vision and Pattern Recognition, Hawaii, USA, 2001, pp. II-31-8, vol.2. 2
8. C. Frueh and A. Zakhor, "Data Processing Algorithms for Generating Textured 3D Building Facade Meshes From Laser Scans and Camera Images", 3D Processing, Visualization and Transmission 2002, Padua, Italy, 2002, pp. 834 - 847
9. N. Haala and C. Brenner, "Generation of 3D city models from airborne laser scanning data", Proc. EARSEL Workshop on LIDAR Remote Sensing on Land and Sea, Tallin, Estonia, 1997, pp.105-112
10. Kim, A. Huertas, and R. Nevatia, "Automatic description of Buildings with complex rooftops from multiple images", Computer Vision and Pattern Recognition, Kauai, 2001, pp. 272-279
11. G. Lowe, "Fitting parametrized three-dimensional models to images", Trans. On pattern analysis and machine intelligence, vol. 13, No. 5, 1991, pp. 441-450
12. H.-G. Maas, "The suitability of airborne laser scanner data for automatic 3D object reconstruction", Third Int'l Workshop on Automatic Extraction of Man-Made Objects, Ascona, Switzerland, 2001
13. I. Stamos and P.E. Allen, "3-D model construction using range and image data." CVPR 2000, Hilton Head Island, 2000, pp.531-6
14. S. Thrun, W. Burgard, and D. Fox, "A real-time algorithm for mobile robot mapping with applications to multi-robot and 3D mapping", ICRA 2000, San Francisco, 2000, vol. 1. 4, pp.321-8