Computers & Graphics 37 (2013) 896–910

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Technical Section

Automatic urban modeling using volumetric reconstruction with surface graph cuts ☆

Ignacio Garcia-Dorado *, Ilke Demir, Daniel G Aliaga

Purdue University, Computer Science, 305 N. University Street, West Lafayette, IN 47907, United States

ARTICLE INFO

Article history:
Received 5 November 2012
Received in revised form
9 July 2013
Accepted 10 July 2013
Available online 1 August 2013

Keywords:
Automatic
Urban
Photo-consistency
Graph cuts
Volumetric reconstruction

ABSTRACT

The demand for 3D city-scale models has been significantly increased due to the proliferation of urban planning, city navigation, and virtual reality applications. We present an approach to automatically reconstruct buildings densely spanning a large urban area. Our method takes as input calibrated aerial images and available GIS meta-data. Our computational pipeline computes a per-building 2.5D volumetric reconstruction by exploiting photo-consistency where it is highly sampled amongst the aerial images. Our building surface graph cut method overcomes errors of occlusion, geometry, and calibration in order to stitch together aerial images and yield a visually coherent texture-mapped result. Our comparisons show similar quality to the manually modeled buildings of Google Earth, and show improvements over naive texture mapping and over space-carving methods. We have tested our algorithms with a 12 sq km area of Boston, MA (USA), using 4667 images (i.e., 280 GB of raw image data) and producing 1785 buildings.

1. Introduction

We present a method for automatic reconstruction of buildings densely spanning a city or portion thereof. The demand for such 3D volumetric content has been significantly increased due to the proliferation of urban planning, city navigation, and virtual reality applications (Fig. 1). Nevertheless, automatic widespread reconstruction of urban areas is still an elusive target. Services, such as Google Earth/Maps, Apple Maps, Bing Maps, and OpenStreetMap have fomented the capture and availability of ubiquitous urban imagery and geographic information system (GIS) style data. Using LIDAR data is one option for city modeling however it still has challenges and is not always available. Ground-level imagery provides high resolution but such images are usually scattered and incomplete. Aerial images provide extensive and uniform coverage of large areas, albeit at lower resolution, and are widely available for most cities. Hence, to reconstruct large urban areas we focus on aerial imagery.

There have been several fundamental approaches for producing urban volumetric reconstructions. In contrast to partial (or facade-level) reconstructions (e.g., Müller et al. [24], Xiao et al. [43]), we seek to automatically create texture-mapped building envelopes spanning a large-portion of a city (i.e., akin to the crowd-sourced created models visible in Google Earth) – such complete models are suitable 3D content for the aforementioned graphics and visualization applications. Inverse procedural modeling approaches pursue generating parameterized 2D and 3D models from observations (e.g., Stava et al. [35], Bokeloh et al. [3], Park et al. [27]), but have not been demonstrated for large-scale urban areas due to the inherent complexity and ambiguity in the inversion process. Relevant volumetric reconstruction methodologies from image-based modeling and computer vision can be loosely divided into (i) space carving and similar techniques (e.g., Kutulakos and Seitz [13], Matusik et al. [22], Montenegro et al. [23], Lazebnik et al. [17], Shalom et al. [32]) and (ii) volumetric graph cuts (e.g., Vogiatzis et al. [41]). All of these methodologies exploit, in some form, photo-consistency, visibility constraints, and smoothness assumptions.

However, for our targeted large areas with high building density and thus a high-level of occlusion, we cannot assume a dense, complete, and un-occluded sampling of all building and ground surfaces. These facts about the input data spawn three important challenges. First, although in a typical aerial capture process each building might be at least partially observed in 25–50 images, parts of each facade might only be seen by a few images (and sometimes none at all). This relatively sparse sampling of the building walls hinders photo-consistency measures. Further, the limited visibility and high-level of occlusion also encumbers the silhouette usage and robust foreground/background segmentation for space carving and hampers the determination of the initial geometry (e.g., visual hull) for volumetric graph cuts. Second, since the captured images of building and ground surfaces may be plagued with the projections of nearby buildings, obtaining occlusion-free projective texture mapping (i.e., texture mapping without neighboring buildings unwillingly appearing on other

☆This article was recommended by A. Shamir.

* Corresponding author. Tel.: +1 765 494 6010; fax: +1 765 494 0739.
E-mail address: igarciad@purdue.edu (I. Garcia-Dorado).

0097-8493/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.cag.2013.07.003

Fig. 1. Urban modeling. A complex urban area (left) is automatically obtained using volumetric reconstruction with surface graph cuts (middle) computed from aerial imagery and GIS-style parcel/building data (right). Our methodology uses photo-consistency to robustly recreate 2.5D building structures and surface graph cuts to assemble seamless and coherent textures despite occlusion, geometry, and calibration errors.

buildings) would require very accurate geometry. Third, obtaining such very accurate geometry is hindered by camera calibration error and by the grazing angle observations of the building facades. Naïve projective and view-dependent texture mapping would produce strong visual discontinuities or would compensate for the inaccuracies by using significant blending/blurring.

Our solution circumvents the aforementioned challenges by exploiting the following inspirations:

- Buildings are, by and large, individual 2.5D structures; thus we assume each successive floor up the building is equal to or contained within the contour of the previous floor.
- Since aerial images mostly sample the roof structures of a building, we exploit photo-consistency only for determining the roof structure; for the building walls, we exploit the 2.5D assumption and stitch together the visual observations using a surface graph-cut based technique (a surface graph cut is a 2D manifold in 3D space that has been stitched together using a solution to the minimum-cost graph-cut problem); our surface graph cut assembles a seamless and visually-coherent texture-mapping of the buildings and ground surfaces despite an imperfect building proxy, projected occlusions, and camera calibration errors.
- To solve the chicken-and-egg dilemma of needing to know the geometry to solve for visibility (and needing to know visibility to solve for geometry), we exploit the assumption of having approximate GIS data (e.g., building outlines) in order to formulate simple building shape estimates which we enhance.

Our approach builds upon voxel occupancy and graph cuts (e.g., Kwatra et al. [14]) to automatically and robustly yield large-scale 3D urban reconstructions. Our largest example includes 1785 reconstructed and texture-mapped buildings spanning more than 12 sq km. Our system pipeline (Fig. 2) takes as input a set of pre-calibrated high-resolution aerial images captured from a multi-camera cluster flying over a city (courtesy of C3Technologies), approximate building outlines extracted from a GIS provider (i.e., OpenStreetMap (OSM)) and rough initial building heights per city zone.

A coarse initial building geometry is subdivided into voxels which are then refined. Improved building outlines, heights, and roof structures are obtained by using a photo-consistency and clustering algorithm. Then, we use surface graph cuts to add the remaining visual details to the building walls and to the ground. The roof structure is sampled by many images. Thus, texture mapping the roof voxels to display additional visual details can be straightforwardly done by selecting the most head-on observations. However, the building walls are sparsely sampled. Hence, in order to create a complete, coherent, and occlusion-free colored appearance, we texture-map wall voxels using the aerial images for which a satisfactory graph cut with the roof and with the adjacent building walls is produced. Further, we solve two other surface graph cut problems in order to provide a smooth visual transition between the building walls and the ground surface as well to produce a top-down high-resolution ground surface image that is free of unwanted projections of building geometry and shadows.

Altogether, our method exploits photo-consistency only where it is highly sampled (thus less susceptible to outliers and noise) and uses a graph-cut based algorithm to stitch together a visually plausible result for the rest of the building surfaces and for the ground surface. Our examples are from a large metropolitan area (i.e., Boston, MA in USA) using a dataset of 4667 aerial images and conservative initial building outlines and height estimates (e.g., often overestimates of 50%). Our comparisons show that our results are significantly better than texturing a space-carving/visual-hull result and similar quality to crowd-sourced manual modeling efforts.

Our main contributions are:

- a robust voxel- and photo-consistency based method for estimating building roof geometry,
- a surface graph cut method to not only stitch textures, but also reduce, or eliminate, artifacts due to incorrect texture overlapping, missing texture fragments, incorrect camera pose, or an inaccurate geometric proxy,
- a graph-cut based method for generating a top-down aerial view of a city free of unwanted building projections appearing on the ground, despite such projections being present in all captured aerial images, and
- a complete automatic framework that generates closed, low polygon count, textured buildings and ground that are ready-to-use in 3D city modeling and computer graphics applications.

2. Related work

In this section, we relate our work to urban modeling approaches in procedural modeling, image-based algorithms,