Graphical Models 74 (2012) 14–28

Contents lists available at SciVerse ScienceDirect

Graphical Models

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

Sorting unorganized photo sets for urban reconstruction

Guowei Wan [a,b,*], Noah Snavely [c], Daniel Cohen-Or [d], Qian Zheng [b], Baoquan Chen [b], Sikun Li [a]

[a] *School of Computer Science, National University of Defense Technology, China*
[b] *Shenzhen Institutes of Advanced Technology, China*
[c] *Department of Computer Science, Cornell University, USA*
[d] *School of Computer Science, Tel Aviv University, Israel*

A R T I C L E I N F O

Article history:
Received 24 August 2011
Received in revised form 24 October 2011
Accepted 4 November 2011
Available online 12 November 2011

Keywords:
Urban modeling
Multi-view stereo
Photo analysis
Structure-from-motion

A B S T R A C T

In spite of advanced acquisition technology, consumer cameras remain an attractive means for capturing 3D data. For reconstructing buildings it is easy to obtain large numbers of photos representing complete, all-around coverage of a building; however, such large photos collections are often unordered and unorganized, with unknown viewpoints. We present a method for reconstructing piecewise planar building models based on a near-linear time process that *sorts* such unorganized collections, quickly creating an image graph, an initial pose for each camera, and a piecewise-planar facade model. Our sorting technique first estimates single-view, piecewise planar geometry from each photo, then merges these single-view models together in an analysis phase that reasons about the global scene geometry. A key contribution of our technique is to perform this reasoning based on a number of typical constraints of buildings. This sorting process results in a piecewise planar model of the scene, a set of good initial camera poses, and a correspondence between photos. This information is useful in itself as an approximate scene model, but also represents a good initialization for structure from motion and multi-view stereo techniques from which refined models can be derived, at greatly reduced computational cost compared to prior techniques.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

Urban scene modeling from imagery is emerging as an important problem in computer graphics and vision, with applications in mapping, virtual tourism, urban planning, architecture, and many other areas. In spite of advanced acquisition technology (e.g., laser scanning), image-based approaches to architectural acquisition and reconstruction are still extremely attractive, due to the ease and low-cost of taking photos. Cameras are portable and ubiquitous, and it is easy to capture images from a wide variety of views to get complete, all-around coverage of a building; we envision large groups of people collaborating to capture such images [1]. Hence, for many buildings we can easily obtain large numbers of photos contributed by everyday photographers. While these collections can represent very rich and comprehensive visual descriptions of a scene, they also tend to be unsorted (or weakly sorted) and unorganized, having been captured by several people from diverse, initially unknown viewpoints. A key challenge is to recover structure latent in such unsorted photo sets.

Recent work in computer vision has established new ways of recovering such structure for architectural scenes, by using image and feature matching techniques to establish correspondences between photos, structure from motion (SfM) techniques for recovering the camera poses [2], and multi-view stereo (MVS) algorithms to compute dense geometry [3,4]. However, these algorithms are often complex and time-consuming. For instance, in order to recover viewpoints, current SfM techniques build up a scene model via an inefficient incremental reconstruction process that can scale poorly with the number of input

* Corresponding author at: School of Computer Science, National University of Defense Technology, China.
E-mail address: gwwan@nudt.edu.cn (G. Wan).

1524-0703/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.gmod.2011.11.001

images. In addition, many urban scenes present key challenges to current approaches, due to symmetries, self-similarity, and insufficient texture evident in many building facades. Hence, given larger and larger photo collections, there is a need for more efficient, robust reconstruction methods for large, unordered collections (see Fig. 1).

In this paper, we present a new approach to reconstruction of architectural scenes that quickly organizes a large photo collection through a process we call *sorting*, which quickly computes a piecewise planar model, estimates a connectivity graph on the image collection, and establishes approximate initial camera poses, as shown in Fig. 2. This information is useful in itself as a simplified building model for graphics applications; moreover, it organizes the photos in a way that can dramatically accelerate further geometric processing. The initial camera poses eliminate much of the work required by SfM, resulting in a significant increase in performance.

Our sorting method is based on a new approach to organizing and reconstructing photos of architectural scenes that first estimates single-view geometry from each image independently, then merges this information to reason about the photo set as a whole. A key contribution is to perform this global reasoning based on a graph-based representation of the geometry, along with a number of new constraints given a piecewise planar assumption, allowing us to quickly derive image matches, approximate global scene geometry, and initial camera poses. While other recent work reasons about camera networks with graphs, we reason directly about *scene structure*. In particular, we treat a building as an ordered sequence of facades (where the number of facades is initially unknown), and we recover this sequence from observations across the entire image set. Our technique first segments out a local, partial facade ordering from each photo, building single-view geometric models. We then reason about how these models fit together in a global analysis stage, in which we compute an ordered set of facades by combining constraints on facade *appearance* and *ordering*, utilizing partial facade orderings observed in each image, as well as building *geometry*, i.e., the constraint that the ordered facades must yield a physically possible building. These three constraints are represented in a *facade graph*, a compact description of the underlying geometry, and we propose a new algorithm that analyzes such a graph to reason about the scene, taking much greater advantage of these constraints than in previous work. Our algorithm is very efficient, produces a useful simplified model, and can significantly accelerate vision pipelines based on SfM and MVS.

In addition to significantly improving speed, our sorting technique is also robust to symmetry and self-similarity by virtue of the proposed geometric constraints. Through analysis of the facade graph, our method can also handle occlusion of facades, is robust to clutter, and can work on scenes with multiple buildings. We integrate our algorithm into a complete dense modeling pipeline, demonstrating results on several large, challenging photo collections of different buildings. These results show that our technique achieves better performance than previous approaches to image-based reconstruction, and can avoid problems due to self-similarity.

As our techniques reason about (planar) facades and their relationships, our sorting algorithm assumes that the input building is approximately piecewise-planar, made up of largely vertical facades. In addition, we make use of lines to distinguish facades, so our technique depends on line segments being evident in the images. While not all buildings fit these assumptions, we demonstrate that our technique works well on a number of real-world buildings, including scenes that are not strictly piecewise planar, as well as collections that give incorrect results using prior methods.

2. Related work

2.1. Efficient image matching

A key ingredient in our work (and any reconstruction pipeline) is establishing correspondence between input images; this is the basis for our "sorting" of images. This correspondence problem is especially challenging for unstructured photo sets, where it is unknown *a priori* which images will match, and where there are often wide baselines. Much recent work has used local image features (e.g., SIFT [5]) and efficient indexing schemes to solve this

Fig. 1. An unorganized photo collection of a building, like the one shown, is sorted to quickly find camera viewpoints, create a piecewise planar model (see Fig. 2), and to facilitate fast multi-view reconstruction.

Download English Version:

https://daneshyari.com/en/article/443049

Download Persian Version:

https://daneshyari.com/article/443049

Daneshyari.com