ISPRS Journal of Photogrammetry and Remote Sensing 98 (2014) 1–18

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Evaluation of feature-based 3-d registration of probabilistic volumetric scenes

Maria I. Restrepo *, Ali O. Ulusoy, Joseph L. Mundy

Brown University, Providence, RI, USA

ARTICLE INFO

Article history:
Received 2 March 2014
Received in revised form 15 September 2014
Accepted 17 September 2014

Keywords:
Probabilistic 3-d modeling
Feature-based 3-d registration

ABSTRACT

Automatic estimation of the world surfaces from aerial images has seen much attention and progress in recent years. Among current modeling technologies, probabilistic volumetric models (PVMs) have evolved as an alternative representation that can learn geometry and appearance in a dense and probabilistic manner. Recent progress, in terms of storage and speed, achieved in the area of volumetric modeling, opens the opportunity to develop new frameworks that make use of the PVM to pursue the ultimate goal of creating an entire map of the earth, where one can reason about the semantics and dynamics of the 3-d world. Aligning 3-d models collected at different time-instances constitutes an important step for successful fusion of large spatio-temporal information. This paper evaluates how effectively probabilistic volumetric models can be aligned using robust feature-matching techniques, while considering different scenarios that reflect the kind of variability observed across aerial video collections from different time instances. More precisely, this work investigates variability in terms of discretization, resolution and sampling density, errors in the camera orientation, and changes in illumination and geographic characteristics. All results are given for large-scale, outdoor sites. In order to facilitate the comparison of the registration performance of PVMs to that of other 3-d reconstruction techniques, the registration pipeline is also carried out using Patch-based Multi-View Stereo (PMVS) algorithm. Registration performance is similar for scenes that have favorable geometry and the appearance characteristics necessary for high quality reconstruction. In scenes containing trees, such as a park, or many buildings, such as a city center, registration performance is significantly more accurate when using the PVM.

1. Introduction

Generating 3-d maps of urban areas from multi-view imagery has seen much attention in recent years, academically and commercially. State of the art frameworks in the area of image-based 3-d modeling or multi-view stereo (MVS) have moved beyond laboratory prototypes to take on the challenge of reconstructing large scale urban models (Hiep et al., 2009; Furukawa and Ponce, 2009; Tola et al., 2011; Pollefeys et al., 2007; Miller et al., 2011). In this domain, illumination and viewing conditions are highly unconstrained and the problem of reconstructing surfaces under the traditional assumption that surfaces exhibit viewpoint independent image intensity is not valid. Among current 3-d modeling technologies, probabilistic volumetric models (PVMs) have evolved as a new type of representation that learns the appearance and geometry of the world surfaces in a dense manner, while encoding the uncertainties caused by occlusions, image noise and limitations of image-based 3-d surface estimation. The advantages of probabilistic learning were initially demonstrated in change detection applications in 2-d images (Pollard and Mundy, 2007; Crispell et al., 2011; Özcanli and Mundy, 2010). More recently, the availability of variable resolution (Crispell et al., 2011) and GPU-based implementations (Miller et al., 2011), has enabled work in the areas of 3-d object classification (Restrepo and Mundy, 2012; Restrepo et al., 2012) and high-resolution surface estimation (Calakli et al., 2012). These works have found advantages when using the dense geometry and appearance stored in the PVM compared to patch-based reconstruction techniques (Furukawa and Ponce, 2009). This work draws inspiration from the advances achieved in prior work using the PVM within a new context: 3-d rigid registration. The variability in the collection characteristics affect the quality of the local 3-d geometry in the models, naturally posing challenges to any application based on the PVM. Therefore, this work presents an in depth analysis with the goal of quantifying how effectively probabilistic volumetric models can be aligned

* Corresponding author.

http://dx.doi.org/10.1016/j.isprsjprs.2014.09.010
0924-2716/

using robust feature-matching techniques considering real-world examples that reflect the kind of variability in illumination and other collection conditions typically observed in aerial video.

Aligning overlapping 3-d models into a common coordinate system is known as 3-d registration. Accurate and efficient solutions to this problem are of interest to a wide variety of fields including computer vision, photogrammetry, robotics, medical imaging and computer graphics. Examples of common applications include: construction of large 3-d maps by combining partial 3-d reconstructions, 3-d instance or landmark recognition, pose estimation of objects and object motion analysis. Prior work using the PVM (Crispell et al., 2011; Restrepo et al., 2012; Calakli et al., 2012) has considered scenes learned from imagery collected on a single day and time of the day as viewed from a helicopter following a ring shaped trajectory. Different models are learned for different instances of these trajectories and no attempt has been performed to merge models containing overlapping geographic areas or models trained from imagery collected on different days. Registering PVM models is an important research problem that can provide means to localize and estimate the pose of specific objects and landmarks, and ultimately help integrating into a single map of the earth, information collected at different instances of time, where one can reason about the state of the static as well as the dynamic information in the models.

Approaches to align 3-d models can be based entirely on geometry (Yamany and Farag, 2002) or on a combination of texture and geometry information (Wu et al., 2008). Purely geometric methods can be used to align 3-d models acquired using active 3-d sensors (Johnson and Hebert, 1999) or to match 2-d frames to 3-d data using lines, curves and other geometric structures (Stamos and Leordeanu, 2003). Studying the scale variability of 3-d models is proposed by Novatnack and Nishino (2008) in order to detect scale-dependent 3-d features. When creating full 3-d models from partial reconstructions, the initial alignment is often considered to be given by sensor information or done manually. Vanden Wyngaerd and Van Gool (2002) propose a patch-based technique to compute these initial positions automatically. Moreover, advances in Structure from Motion (SfM) and 3-d modeling have generated interest in the alignment of 3-d scenes using large collections of photos with a wide baseline (Snavely et al., 2006), the alignment of 3-d points obtained from SfM to range data (Zhao et al., 2005) and more recently on temporal analysis of 3-d urban scenes (Ulusoy and Mundy, 2014; Schindler and Dellaert, 2010; Taneja et al., 2011).

In the context of 3-d to 3-d registration, techniques used to recover the transformation between 3-d models often rely on matching geometric features, also know as feature-based registration. While the matching performance of many 3-d shape descriptors has been studied in mesh or point cloud data (from range sensors, or computer generated meshes) it remains unclear whether their descriptiveness and robustness to noise successfully extends to the geometry learned using MVS systems and in particular to the diffuse surface probability distributions of the PVM. In that regard, Restrepo and Mundy (2012) investigate the robustness of several shape descriptors for object categorization using the PVM, where encouraging recognition rates are observed for the Fast Point Feature Histogram (FPFH) (Rusu et al., 2009) and the Signatures of Oriented Histograms (SHOT) (Tombari et al., 2010). Motivated by those results, this paper evaluates the robustness of the SHOT and FPFH descriptors for model registration using feature-matching techniques. Furthermore, the quality of the underlying geometry, and therefore the effectiveness of the descriptors for matching purposes, is affected by variations in the conditions of the data collection. A major contribution of this work is to evaluate and compare the quality of feature-based registration between PVM and PMVS models under different scenarios that reflect the kind of variability observed across collections from different times instances, including model discretization, resolution, sampling density, errors in the camera orientation, and changes illumination and in geographic characteristics.

Feature-based alignment can also be achieved using robust image features and SfM techniques as it is often done in the area of 2-d to 3-d pose estimation (Li et al., 2012; Hao et al., 2013). However, this work focuses on pure 3-d features for few reasons. First, using only the 3-d geometry in the models allows the system to save storage and forget the images used during reconstruction. This advantage is particularly relevant if models are learned using images collected during large periods of time. Second, using shape features sets the stage for more general registration systems that bring into alignment other 3-d modalities e.g. LiDAR, to further enrich the modeling of urban scenes. Finally, shape features relate directly to the geometry of objects, which can lead to a better semantic understanding of 3-d scenes.

Correspondence-based techniques can solve the registration problem without the need of an initial guess. However, the accuracy of the result is limited by presence of large number of false matches and noise in the measurements, and can typically be improved using local iterative algorithms. The Iterative Closest Point algorithm (ICP) is used in this work to provide further intuition as to under which conditions, feature-based registration techniques can successfully find an good initial alignment i.e. such that local registration algorithms converge to precise estimates. Fig. 1 presents an overview of the registration pipeline used here for evaluation. First, the probabilistic models are learned from data collected under different illumination conditions and from different viewing directions. The resulting models lie in different coordinate systems. Then, rough alignment of the models is achieved using feature-based matching. Finally, a more refined alignment is achieved using ICP.

2. Background

This section briefly summarizes the related work needed to understand the study presented in this paper. A more in-depth treatment of the subjects can be found in Restrepo (2013).

2.1. Probabilistic Volumetric Model: PVM

The probabilistic volumetric models used in this study were first proposed by Pollard and Mundy (2007). The framework learns surface probabilities and the appearances of voxels using an online Bayesian learning algorithm along projection rays. Specifically, the probability of a voxel being part of a surface is updated in an online fashion with the intensity observed in a pixel associated with a corresponding projection ray. Surface probability increases if the appearance model at the voxel explains the intensity seen in the images better than any other voxel along the ray. The appearance at a voxel is modeled with a Gaussian mixture distribution. For details on related update equations please refer to Pollard and Mundy (2007)'s original work.

The model just described is implemented in a regular grid and it does not scale up to large scenes due to its cubic space complexity. Crispell et al. (2011) address these limitations with a continuous formulation of Pollard and Mundy's method implemented in an octree. Moreover, Miller et al. (2011) propose a GPU implementation that is capable of learning large, high resolution volumetric models efficiently, where one pass of the online update can be processed at approximately one frame per second for HD resolution.