ISPRS Journal of Photogrammetry and Remote Sensing 142 (2018) 328–341

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Scalable non-rigid registration for multi-view stereo data

Gianpaolo Palma[a,*], Tamy Boubekeur[b], Fabio Ganovelli[a], Paolo Cignoni[a]

[a] Visual Computing Lab – ISTI CNR, Pisa, Italy
[b] LTCI, Telecom ParisTech, Paris-Saclay University, Paris, France

ARTICLE INFO

Keywords:
Non-rigid registration
Multi-view-stereo
Low-frequency deformation
Scalable implementation

ABSTRACT

We propose a new non-rigid registration method for large 3D meshes from Multi-View Stereo (MVS) reconstruction characterized by low-frequency shape deformations induced by several factors, such as low sensor quality and irregular sampling object coverage. Starting from a reference model to which we want to align a new 3D mesh, our method starts by decomposing it in patches using a Lloyd clustering before running an ICP local registration for each patch. Then, we improve the alignment using few geometric constraints and finally, we build a global deformation function that blends the estimated per-patch transformations. This function is structured on top of a deformation graph derived from the dual graph of the clustering. Our algorithm is iterated until convergence, increasing progressively the number of patches in the clustering to capture smaller deformations. The method comes with a scalable multicore implementation that enables, for the first time, the alignment of meshes made of tens of millions of triangles in a few minutes. We report extensive experiments of our algorithm on several dense Multi-View Stereo models, using a 3D scan or another MVS reconstruction as reference. Beyond MVS data, we also applied our algorithm to different scenarios, exhibiting more complex and larger deformations, such as 3D motion capture dataset or 3D scans of dynamic objects. The good alignment results obtained for both datasets highlights the efficiency and the flexibility of our approach.

1. Introduction

The generation of 3D surfaces from measured data is a very important task in the acquisition and computation of complete digital representations of real-world objects. Among the various 3D scanning technologies, Multi-View Stereo (MVS) reconstruction from images appears as a very cost-effective solution: the ubiquitous and wide availability of cameras gives everyone the possibility of harvesting, in a short time and with inexpensive hardware, many images to use for the 3D MVS reconstruction of the world around us. Furthermore, the rise and the consolidation of the community photo collections have increased considerably the amount of data available for this purpose (Goesele et al., 2007; Agarwal et al., 2009; Heinly et al., 2015). The 3D models obtained with these technologies can be used for different applications, among which multimodal capture is becoming more and more frequent. In this context, the MVS data can be used for two different purposes: the 3D model completion or the temporal environment monitoring. In the first case, the goal is to enrich an existing 3D model obtaining a complete sampling (for example to integrate and complete a high-quality 3D scan of a monument with the missing parts that are easier to acquire by photogrammetry with a drone). In the second case, the goal is to monitor the temporal shape evolution of an environment or an object automatically comparing a pair of 3D models acquired at different points in time. For both applications, a fundamental step resides in the registration of the computed MVS mesh to a reference model (e.g., a high-quality 3D model, a laser scan or another MVS reconstruction). Two separate issues hinder this goal. The first one is the estimation of the unknown scale factor of the new MVS mesh with respect to the reference model. A solution for this issue was proposed in Mellado et al. (2016) for general 3D models and in Persad and Armenakis (2017) when the input can be approximated with a height map. The second issue is the low frequency deformation introduced in the model by the Structure-From-Motion (SfM) step of the MVS reconstruction (see the input MVS model in Fig. 1) and related to several factors: (i) the low sensor quality; (ii) the sampling acquisition coverage of the objects in term of overlap among the images; (iii) the spatial distribution of the views; (iv) the parameters of the involved algorithms that are not able to remove all the optical lens distortion or to detect a sufficient number of features to constrain the bundle adjustment step. Even if some of these problems can be controlled in the case of an ad-hoc photographic campaign, for example taking care of the sampling acquisition coverage or using expensive devices like a total station theodolite to acquire a number of Ground Control Points, the situation is worse in the case of community photo collections where the

* Corresponding author.
E-mail address: gianpaolo.palma@isti.cnr.it (G. Palma).

https://doi.org/10.1016/j.isprsjprs.2018.06.012
Received 12 February 2018; Received in revised form 4 May 2018; Accepted 18 June 2018
0924-2716/

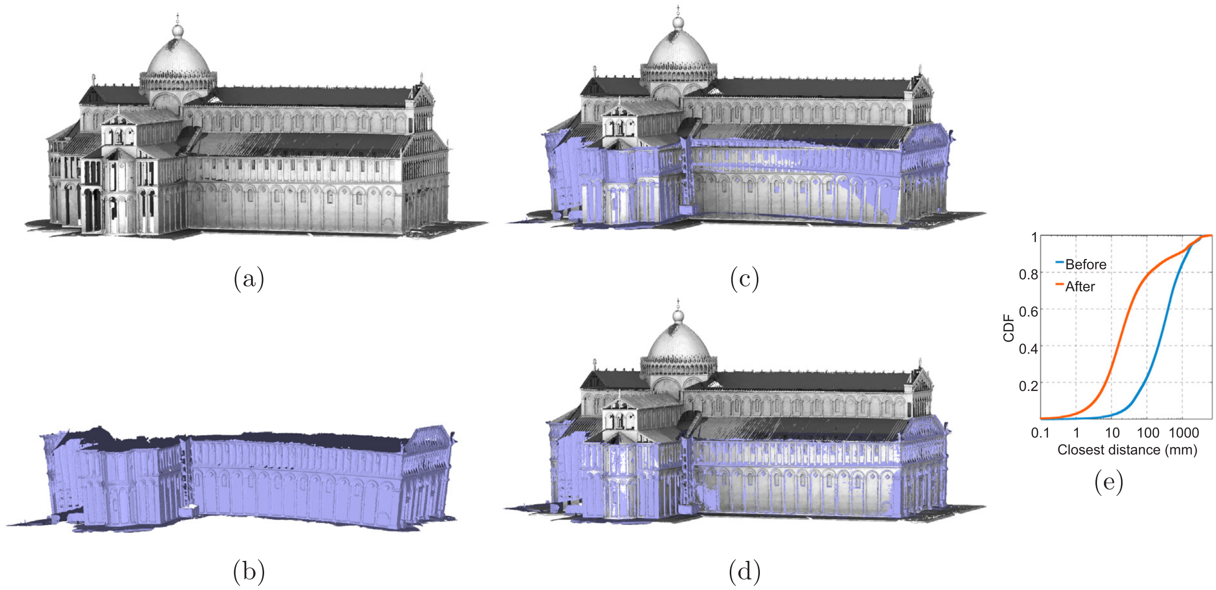

Fig. 1. Our algorithm deforms a 3D model obtained with multi-view stereo (MVS) methods to align it over a reference mesh. (a) Input reference mesh by 3D scanning. (b) Input deformable mesh by MVS reconstruction. (c) Initial rigid alignment. (d) Final alignment computed by our algorithm. (e) Cumulative Distribution Function (CDF) of the vertex-to-mesh distance between the deformed mesh and the reference one (blue before and orange after the alignment). Our algorithm corrects the bending introduced by the MVS reconstruction along the nave of the cathedral, preserving the significant geometric changes like the scaffolding on the side of the building. This is also confirmed by the better CDF of the closest distance after the alignment. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

acquisition is incidental and the photos are acquired with different cameras.

In this paper, we propose an iterative and hierarchical non-rigid registration method to align a large MVS mesh to a reference 3D model while removing the deformation introduced by the MVS algorithm. The final goal is to correct the deformation in a relative manner since the reference model continues to have its own distortions. This relative alignment can be very useful for the applications that can take advantage of an acquisition with inexpensive hardware, like a camera, where the relative deformation of one model to the previously acquired one permits or to integrate the data of the two models or to detect the changed regions in a more robust way. Starting from a rough alignment to the reference mesh, at each iteration, our algorithm segments the model into patches using a Lloyd clustering. The resulting segmentation is used to build a deformation graph that allows transferring the affine transformation computed for each node of the graph onto the original model. The affine transformations are computed independently for each patch with an Iterative Closest Point (ICP) procedure, using the geometry of the adjacent patches to constrain the local stiffness. This local ICP procedure makes the algorithm easier to parallelize on multicore architectures, allowing, for the first time, the alignment of meshes with tens of millions of triangles in few minutes. When the ICP converges to a local minimum, we improve the estimation by regularizing the local consistency of the transformation from the adjacent nodes. All along the iterations, we increase the number of patches of the Lloyd clustering to progressively fit smaller scale deformations.

Our main contribution is a non-rigid registration algorithm that allows processing very large and complex MVS meshes. We do not make any assumptions about the type of the input data and the amount of deformation between the models, as opposed to the state-of-the-art non-rigid registration algorithms that manage specific input, like range scan, taking advantage from their implicit 2D parameterization (for example Li et al., 2008; Chang and Zwicker, 2009), or range video, taking advantage of the temporal coherence of the data (for example Li et al., 2013; Zhou et al., 2013). These assumptions prevent the trivial application of these solutions to our kind of input, and their adaptation to our input is not straightforward. Other state-of-the-art solutions are based on complex global non-linear energy minimization (Cagniart et al., 2010), which do not scale with the size of the input models. On the contrary, our solution is based on a localized ICP and local updates of the deformation model, making the algorithm easier to run in parallel and scalable to large 3D meshes. In particular, our algorithm is robust against all usual defects exhibited by a typical MVS mesh such as noise, missing geometry, irregular triangulation or irregular density. It can also handle multi-scale input with very different level-of-details in the same mesh and between the target and the deformed mesh. Finally, our method is also robust to the presence of geometric changes, like pieces of new geometry that did not exist in the reference mesh, preserving these changes (Fig. 11). This property is key for the detection of changes in evolving 3D data where MVS techniques offer a practical cost-effective solution. We tested the algorithm with different real datasets using reference models of different quality, such as a high-quality 3D scanned model, a single raw LIDAR scan or another MVS model. Furthermore, we tested different scenarios that exhibit more complex and larger deformations, such as 3D motion capture datasets or dynamic object scans, showing that, although specifically designed for MVS data, our method achieves comparable results with a state-of-the-art algorithm on a broader set of application scenarios.

2. Related work

The deformations in an MVS reconstruction are mainly due to the SfM step that introduces drifting effects due to different reasons (low sensor quality, not good view sampling of the scene, the parameters of the involved algorithms). Cohen et al. (2012) propose a new SfM formulation to solve this problem in the case of architectural scenes with symmetric or repeated structures. Although this SfM method returns a more natural coordinate system, reducing the final deformation, the input requirements prevent its application to more general scenes. On the contrary, we formulate the problems as a non-rigid registration using another 3D mesh as reference.

The literature of the 3D registration problem is very wide and in this section, we analyze the approaches more closely related to the proposed method. Our starting point is the Iterative Closest Point (Besl and