

Harvesting Dynamic 3D Worlds from Commodity Sensor Clouds

T. Boubekeur¹, P. Cignoni², E. Eisemann³, M. Goesele⁴, R. Klein⁵, S. Roth⁴, M. Weinmann⁵, and M. Wimmer⁶

¹Telecom ParisTech, Paris, France

²CNR-ISTI, Pisa, Italy

³Technical University Delft, Delft, The Netherlands

⁴Technische Universität Darmstadt, Darmstadt, Germany

⁵Universität Bonn, Bonn, Germany

⁶Technische Universität Wien, Wien, Austria

Abstract

The EU FP7 FET-Open project “Harvest4D: Harvesting Dynamic 3D Worlds from Commodity Sensor Clouds” deals with the acquisition, processing, and display of dynamic 3D data. Technological progress is offering us a wide-spread availability of sensing devices that deliver different data streams, which can be easily deployed in the real world and produce streams of sampled data with increased density and easier iteration of the sampling process. These data need to be processed and displayed in a new way. The Harvest4D project proposes a radical change in acquisition and processing technology: instead of a goal-driven acquisition that determines the devices and sensors, its methods let the sensors and resulting available data determine the acquisition process. A variety of challenging problems need to be solved: huge data amounts, different modalities, varying scales, dynamic, noisy and colorful data. This short contribution presents a selection of the many scientific results produced by Harvest4D. We will focus on those results that could bring a major impact to the Cultural Heritage domain, namely facilitating the acquisition of the sampled data or providing advanced visual analysis capabilities.

Categories and Subject Descriptors (according to ACM CCS): Computer Graphics [Computing methodologies]: Shape Modeling—

1. Introduction

The EU FP7 FET-Open project Harvest4D deals with the acquisition, processing, and display of dynamic 3D data. The traditional acquisition process for visual models of 3D worlds requires:

- planning a specific scanning campaign,
- carefully selecting the (often costly) acquisition devices,
- performing the on-site acquisition at the required resolution,
- post-processing the acquired data to produce a beautified triangulated and textured model.

However, technological progress led to the wide-spread availability of sensing devices that deliver diverse data streams, that can be easily deployed in the real world, and that produce data with increased density and easier iteration of the sampling process. These data need to be processed and displayed in a new way.

To exploit this development, the Harvest4D project proposes a radical change in acquisition and processing technology: instead of a goal-driven acquisition that determines the devices and sensors, its methods let the sensors and resulting available data determine the acquisition process. A variety of challenging problems arise:

huge data amounts, different modalities, varying scales, dynamic, noisy and colorful data.

Harvest4D involves six partners from five EU member states: TU Wien (AT, coordinator), Universität Bonn (DE), Telecom ParisTech (FR), TU Darmstadt (DE), CNR-ISTI (IT), and Technical University Delft (NL).

Harvest4D produced excellent results, scoring as one of the best among recent FET projects, with 30 papers published in top-ranked venues (e.g., more than 6 ACM Siggraph/TOG and 10 CGF/Eurographics publications). The full list of scientific results is available on the project webpage at <http://harvest4d.org>.

This short contribution presents a selection of the many scientific results produced by Harvest4D, with a focus on those impacting Cultural Heritage (CH). One Harvest4D application scenario is 3D sampling/documentation of a large evolving infrastructure; an immediate example is the management of a major archaeological site, which encompasses all phases of the digitization and digital access pipeline. In this specific domain, we need new technologies for *advanced sampling* of evolving scenarios (e.g., the evolving progress of an archaeological excavation) and for *processing* the huge quan-

tity of 3D samples produced. Once massive high-resolution 3D models are in place, the challenge is how to provide ubiquitous and easy interactive access to those data. The Harvest4D project thus also deals with new *visualization and interaction* paradigms, e.g., visualization of large 3D data over the web.

2. Advanced Sampling Techniques

Our novel Harvest4D technologies facilitate revolutionary 3D model acquisitions, especially w.r.t. CH. Following the Harvest4D idea that sensor devices define the acquisition process, our algorithms are robust, flexible and pose only little demands to the user.

Particularly, we provide the *Multi-View Environment* (MVE) [FLM*15], a step-based, end-to-end reconstruction pipeline. Initially, Structure from Motion (SfM) is run on uncontrolled image sets to register cameras and compute a sparse scene point cloud. Then, Multi View Stereo (MVS) is executed to output a dense depth map per registered camera. Finally, Surface Reconstruction (SR) is performed on a dense and possibly huge point cloud from the depth maps to create a colored mesh (Figure 1).

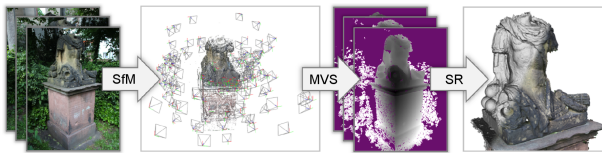


Figure 1: Complete reconstruction of a memorial using MVE.

With the Harvest4D goals in mind, we devised an innovative SR algorithm, Floating Scale Surface Reconstruction (FSSR) [FG14]. FSSR excels at processing multi-scale, oriented, uncontrolled, redundant and potentially very noisy point clouds. Most importantly, our method takes into account the inherent and varying scales of 3D samples. FSSR produces globally consistent models for input data that has fine details of interesting scene parts and low-resolution data for the rest. It draws upon a simple yet effective mathematical formulation to construct a global, implicit function as the weighted sum of compactly supported basis functions. The function has a spatially continuous floating scale to represent the input data detail level appropriately. Finally, FSSR requires no preprocessing, is easy to implement, parameter free and highly scalable owing to no global operations.

According to the Harvest4D acquisition paradigm, the MVE pipeline is easy to use. First, reconstruction requires almost no manual intervention as, in contrast to previous approaches, all steps automatically control their behavior in a data-driven way depending on given scene conditions. Second, MVE’s scene-specific way of processing supports a wide scene variety, handling strongly varying detail levels, different illumination conditions, occlusions, etc. Support for common data formats allows for easy extra processing, exchange of steps and different data types, e.g., processing of aligned range scans by FSSR, skipping SfM and MVS. Third, allowing free capturing and use of cheap, mobile sensor devices, MVE keeps costs very low.

Another key aspect of Harvest4D is the *integration of different modalities*, which includes a fully automated pipeline for registering images to an untextured 3D model of a real-world scene [PR15]. Since neither the texture of the 3D model nor the lighting conditions at capture time are known, we rely on a newly proposed gradient rendering technique, called Average Shading Gradients, to deal with these challenges. Our pipeline proceeds by first computing coarse pose estimates that later get refined and finally verified. In the CH context, our method allows aligning contemporary photographs as well as historic depictions of a scene (Figure 2).



Figure 2: Our image-to-geometry registration pipeline copes with photographs as well as historic depictions of a scene [PR15].

3. Advanced Processing of Sampled Data

Processing huge, potentially time-varying datasets is not a simple task. To deal with this challenge, Harvest4D offers a new solution for *accurate real-time simplification* [LB15], based on the new concept of Morton Integrals. By summing up quadric error metric matrices along Morton-ordered surface samples, we extract concurrently the nodes of an adaptive cut in the so-defined implicit hierarchy, and optimize all simplified vertices in parallel. This approach is inspired by integral images and exploits recent advances in high-performance spatial hierarchy construction and traversal. As a result, the GPU implementation can down-sample a mesh (resp. a point cloud) made of several millions of vertices in real time, at the speed data is generated, while providing better quality than uniform simplification and preserving important salient features (Figure 3).

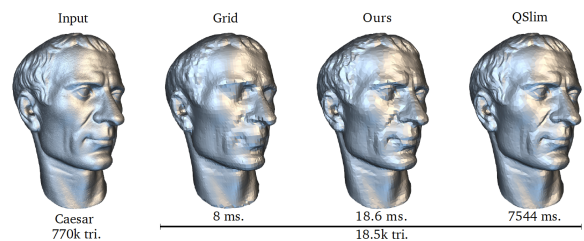


Figure 3: An example of the results obtained in real time using the new geometric simplification algorithm [LB15].

Another important issue is *detecting geometric changes* between different samplings of the same scene, performed by multiple acquisitions occurring at different times. This is a critical operation for all systems requiring a precise segmentation between changing and non-changing regions. Unfortunately, typical 3D scanning setups cannot provide a one-to-one mapping between measured samples in static regions: in particular, both extrinsic and intrinsic sen-

sensor parameters may vary over time, while sensor noise and outliers additionally corrupt the data. Harvest4D introduces a multi-scale approach to robustly tackle these issues [PCBS15]. Starting from two point clouds, first outliers are removed using a probabilistic operator. Then, the actual change is detected using the implicit surface defined by the point clouds under a Growing Least Squares reconstruction that, compared to the classical proximity measure, offers a more robust change/no-change characterization near the temporal intersection of the scans and in the areas exhibiting different sampling density and direction. The resulting classification is enhanced with a spatial reasoning step to solve critical geometric configurations that are common in man-made environments.

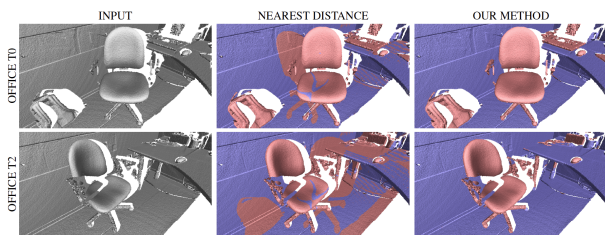


Figure 4: An example of the results obtained with the new change detection algorithm [PCBS15].

Finally, an efficient storage of the huge datasets typically produced in CH applications is challenging and leads to the need for data compression. A *fast, multi-modal point cloud compression* approach that outperforms previous techniques regarding storage requirements as well as computational cost has been developed in Harvest4D [GK15]. By encoding geometric details (in terms of height maps) as well as colors parameterized over a locally planar approximation of the scene geometry in respective texture representations and applying state-of-the-art image compression, high compression ratios can be achieved at rather low computational costs while maintaining an accurate representation of the scene (see Figure 5).

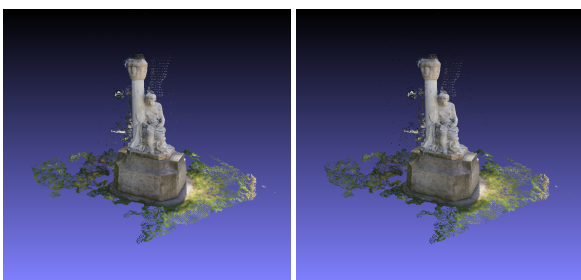


Figure 5: Original point cloud and point cloud obtained after decompressing the compressed data [GK15].

4. Advanced Visualization

Providing interactive visualization for the huge datasets produced in the CH domain is not an easy task. We have either digitization technologies able to produce tens of millions of samples/triangles

even in the case of small or medium scale artworks; or, in other application cases, we could be interested in sampling very large-scale subjects, up to the size of an archaeological site or an entire historical city. The issues related to the interactive visualization (either local or remote) and the need of an effective approach to interact with and navigate complex spaces are a key focus in Harvest4D.

The consolidated acquisition and reconstruction approach (range maps registration, global reconstruction and texturing) could become impractical with increasing data sizes, making it difficult to obtain visualizations of the data in reasonable time. We therefore solve the problem of texturing and reconstruction locally [APS*14], using depth maps generated for each input image. In a continuation of this work, the progress of GPU technology and raytracing solutions allows us to skip even more pre-processing requirements, going to *raytracing of multiple depth maps* directly [APW16]. In a preprocessing phase, we first generate high-resolution textured depth maps by rendering the input points from image cameras and then perform a graph-cut based optimization to assign a small subset of these points to the images. At runtime, we use the resulting point-to-image assignments (1) to identify for each view ray which depth map contains the closest ray-surface intersection and (2) to efficiently compute this intersection point. The resulting algorithm accelerates both the texturing and the rendering of the depth maps by an order of magnitude (see Figure 6), and provides seamless texturing of the dataset.

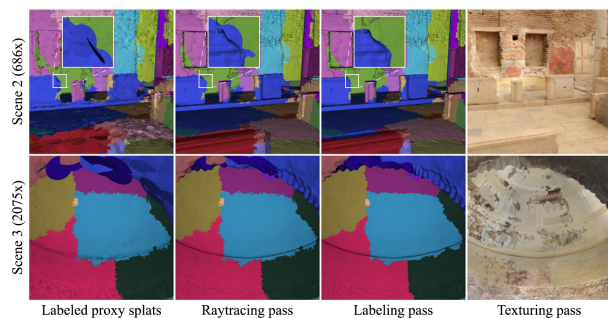


Figure 6: An example of the results obtained with the new raytracing solution [APW16].

Another alternative approach is based on the conversion from sampled points or surfaces to a *compressed voxel dataset*. Voxel representations, coupled with raytracing rendering approaches have been used since two decades. Complexity resides in the need of providing very high resolution and, usually, dealing with uneven densities (as in most sampled datasets). Recently, directed acyclic graphs (DAGs) were successfully used for compressing sparse voxel scenes, but they are restricted to a single bit of (geometry) information per voxel. In Harvest4D, we present a method to compress arbitrary data, such as colors, normals, or reflectance information [DKB*16]. By decoupling geometry and voxel data via a novel mapping scheme, it is possible to apply the DAG principle to encode the topology, while using a palette-based compression for the voxel attributes, leading to a drastic memory reduction. This method outperforms existing state-of-the-art techniques and is well-suited for GPU architectures, achieving real-time performance

on commodity hardware for colored scenes with up to 17 hierarchical levels ($128K^3$ voxel resolution), which are stored fully in core (see Figure 7).



Figure 7: Compressed voxelized scene at different levels of detail, rendered in real time using raytracing only [DKB*16].

The remote visualization and navigation of 3D data directly inside a web browser is becoming feasible due to recent efforts in standardizing the components for 3D rendering. A method for the easy remote navigation of complex archaeological 3D environments [CDS15], represented by multi-resolution triangle meshes, has been implemented on top of the 3DHOP platform [PCD*15]. This system supports two intuitive navigation modes (see Figure 8): the user can explore the model from the top (bird's eye mode, left image) or can walk inside the environment in a walk-through fashion (first-person mode, right image). Furthermore, we also provide a 3D web-based viewer, Potree, that can visualize huge point clouds directly in high quality [SW15], and allows extracting profiles, performing measurements, etc.

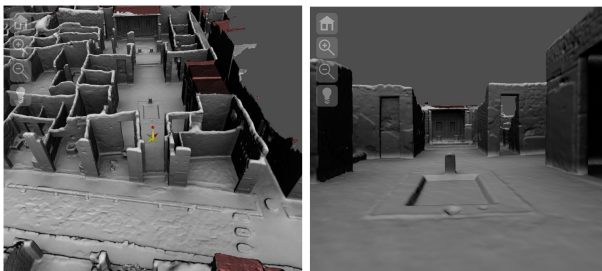


Figure 8: Web-based visual navigation over Pompeii's houses: switching from bird's eye to first-person mode preserves the user position and orientation.

5. Conclusions

Harvest4D has significantly contributed to technological progress by proposing a radical change in acquisition and processing pipelines focusing on the efficient management of huge amounts of 3D acquired data with different modalities, in varying scales, and considering dynamic, noisy and appearance rich environments.

In this paper, we described several important contributions Harvest4D made to the Cultural Heritage (CH) domain. For example, the proposed technologies can be of paramount importance in 3D sampling/documentation of large evolving environments like the management of a major archaeological site, a problem that encompasses all phases of the digitization and digital access pipeline. In this specific domain, the project provides new technologies for

sampling evolving scenarios (e.g., the evolving progress of an archaeological excavation) and for processing the huge quantity of 3D samples produced. Moreover, for the resulting massive high-resolution 3D models, the project has provided new visualization and interaction paradigms and tools for visualization of large 3D data over the web. Please visit <http://harvest4d.org> for more showcases, open source software and datasets from the project.



References

- [APS*14] ARIKAN M., PREINER R., SCHEIBLAUER C., JESCHKE S., WIMMER M.: Large-scale point-cloud visualization through localized textured surface reconstruction. *IEEE Transactions on Visualization and Computer Graphics* 20, 9 (2014), 1280–1292. 3
- [APW16] ARIKAN M., PREINER R., WIMMER M.: Multi-depth-map raytracing for efficient large-scene reconstruction. *IEEE Transactions on Visualization and Computer Graphics* 22, 2 (Feb 2016), 1127–1137. doi:10.1109/TVCG.2015.2430333. 3
- [CDS15] CALLIERI M., DELLEPIANE M., SCOPIGNO R.: Remote visualization and navigation of 3d models of archeological sites. In *ISPRS Archives, Proceedings of 3D-ARCH Conference* (2015), vol. XL-5/W4, ISPRS, pp. 147–154. URL: <http://vcg.isti.cnr.it/Publications/2015/CDS15>. 4
- [DKB*16] DADO B., KOL T. R., BAUSZAT P., THIERY J.-M., EISEMANN E.: Geometry and attribute compression for voxel scenes. *Computer Graphics Forum (Proc. Eurographics)* 35, 2 (may 2016). URL: <http://graphics.tudelft.nl/Publications-new/2016/DKBTE16>. 3, 4
- [FG14] FUHRMANN S., GOESELE M.: Floating scale surface reconstruction. *ACM Trans. Graph.* 33, 4 (July 2014), 46:1–46:11. doi:10.1145/2601097.2601163. 2
- [FLM*15] FUHRMANN S., LANGGUTH F., MOEHRLE N., WAECHTER M., GOESELE M.: Mve-an image-based reconstruction environment. *Comput. Graph.* 53, PA (Dec. 2015), 44–53. doi:10.1016/j.cag.2015.09.003. 2
- [GK15] GOLLA T., KLEIN R.: Real-time point cloud compression. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (2015). 3
- [LB15] LEGRAND H., BOUBEKEUR T.: Morton integrals for high speed geometry simplification. In *Proceedings of the 7th Conference on High-Performance Graphics* (New York, NY, USA, 2015), HPG '15, ACM, pp. 105–112. doi:10.1145/2790060.2790071. 2
- [PCBS15] PALMA G., CIGNONI P., BOUBEKEUR T., SCOPIGNO R.: Detection of geometric temporal changes in point clouds. *Computer Graphics Forum* (2015). In press. URL: <http://vcg.isti.cnr.it/Publications/2016/PCBS16>. 3
- [PCD*15] POTENZIANI M., CALLIERI M., DELLEPIANE M., CORSINI M., PONCHIO F., SCOPIGNO R.: 3DHOP: 3D Heritage Online Presenter. *Computers and Graphics* 52 (2015), 129 – 141. doi:http://dx.doi.org/10.1016/j.cag.2015.07.001. 4
- [PR15] PLÖTZ T., ROTH S.: Registering images to untextured geometry using average shading gradients. In *IEEE International Conference on Computer Vision (ICCV)* (Santiago, Chile, Dec. 2015), pp. 2030–2038. doi:10.1109/ICCV.2015.235. 2
- [SW15] SCHÜTZ M., WIMMER M.: High-quality point based rendering using fast single pass interpolation. In *Proceedings of Digital Heritage 2015 Short Papers* (Sept. 2015), pp. 369–372. URL: <https://www.cg.tuwien.ac.at/research/publications/2015/SCHUETZ-2015-HQP/>. 4