HARVEST4D

HARVESTING DYNAMIC 3D WORLDS FROM COMMODITY SENSOR CLOUDS

Project Periodic Report Summary

Month 12

Date:	24.7.2014
Grant Agreement number:	EU 323567
Project acronym:	HARVEST4D
Project title:	Harvesting Dynamic 3D Worlds from Commodity Sensor Clouds

TABLE OF CONTENTS

1 PUBLISHABLE SUMMARY

1.1 PROJECT CONTEXT AND OBJECTIVES

The current acquisition pipeline for visual models of 3D worlds is based on a paradigm of planning a goal-oriented acquisition - sampling on site - processing. The digital model of an artifact (an object, a building, up to an entire city) is produced by planning a specific scanning campaign, carefully selecting the (often costly) acquisition devices, performing the on-site acquisition at the required resolution and then post-processing the acquired data to produce a beautified triangulated and textured model. However, in the future we will be faced with the ubiquitous availability of sensing devices that deliver different data streams that need to be processed and displayed in a new way, for example smartphones, commodity stereo cameras, cheap aerial data acquisition devices, etc.

We therefore propose a radical paradigm change in acquisition and processing technology: instead of a goal-driven acquisition that determines the devices and sensors, we let the sensors and resulting available data determine the acquisition process. Data acquisition might become incidental to other tasks that devices/people to which sensors are attached carry out. A variety of challenging problems need to be solved to exploit this huge amount of data, including: dealing with continuous streams of time-dependent data, finding means of integrating data from different sensors and modalities, detecting changes in data sets to create 4D models, harvesting data to go beyond simple 3D geometry, and researching new paradigms for interactive inspection capabilities with 4D datasets. In this proposal, we envision solutions to these challenges, paving the way for affordable and innovative uses of information technology in an evolving world sampled by ubiquitous visual sensors.

1.2 PROGRESS AND MAIN RESULTS

1.2.1 Introduction and Highlights

The project is structured around 6 main research objectives, which together address the challenge of incidental capture of dynamic 3D data. Work packages 3-8 deal with these research objectives, while work package 9 addresses the integration of the individual results. In **WP3: Flash Geometry Processing**, GPU-based geometric operators are designed as basis for WP4-WP8. **WP4: Integration of Modalities**, deals with algorithms to integrate different modalities like images, video and range scanner input into a common coordinate frame. This is usually the first stage that input data has to take so that subsequent stages can operate on them. In **WP5: Multi-Scale Scene Processing**, we lift data input from WP4 to a more efficient representation level based on multi-scale processing, and separating noise from data. **WP6: Detecting Changes** tries to detect and separate actual changes in the world from noise based on data from WP5. **WP7: Reflectance Reconstruction** deals with the challenge of reconstructing more high-level reflectance information. In **WP8: Visual Interaction Paradigms**, we design new visualization algorithms for the efficient visual presentation of the data provided by WP4-WP7, together with new interaction mechanisms for these data. Research concentrates on challenges concerning both large data visualization and interaction.

In the first year of the project, the specific goals were to advance the theoretical foundations and to start the development of algorithms for each of these research objectives, resulting in both scientific publications and actual prototypes. Accordingly, all partners have made **progress to achieve the individual work package objectives**. This has already led to **more than 25 peer-reviewed scientific publications**. In particular, 3 papers appear at the **ACM SIGGRAPH 2014** conference, the most highly ranked venue in the field of the project.

In addition, partners have already **worked together** to define interfaces and experiment with pipelines **to pave the way for the future integration** of the individual components of the project. To facilitate this, several **test data sets were created** in a fashion that mimics incidental data capture.

Finally, we believe we have discovered a **potential new unifying representation** of incidentally captured data sets, so-called Gaussian Mixture Models for points. This new continuous representation of data sets formed the basis of two successful publications in the project.

In the following, we summarize the individual results achieved so far in each scientific work package (WP3-WP8), as well as progress in integration (WP9).

1.2.2 WP3: Flash Geometry Processing

Generating consolidated 3D models of the acquired world requires, at every single stage, the execution of geometry processing algorithms that improve, adapt and transform the 3D data either to a particular application scenario using the acquired data, or to a particular integration modality which stores efficiently the cumulative sum of all acquired geometry information. Examples of geometric operators include simplification (see Figure 1), resampling, meshing or filtering.

During the first year of the project, this work package has provided the Harvest4D capture pipeline with efficient geometric primitives for a large collection of applications, allowing approximating, resampling or analyzing the geometric data instantaneously. Here, the key idea was that when one deals with a large number of continuous streams of data, long preprocessing is not an option for at least two reasons: firstly, the mandatory computational capacity may not be available or large enough to perform a given processing stage before the next batch of data appears on the stream. Secondly, pretty much any geometric operator loses a bit of the original information when seeking for its own optimal target.

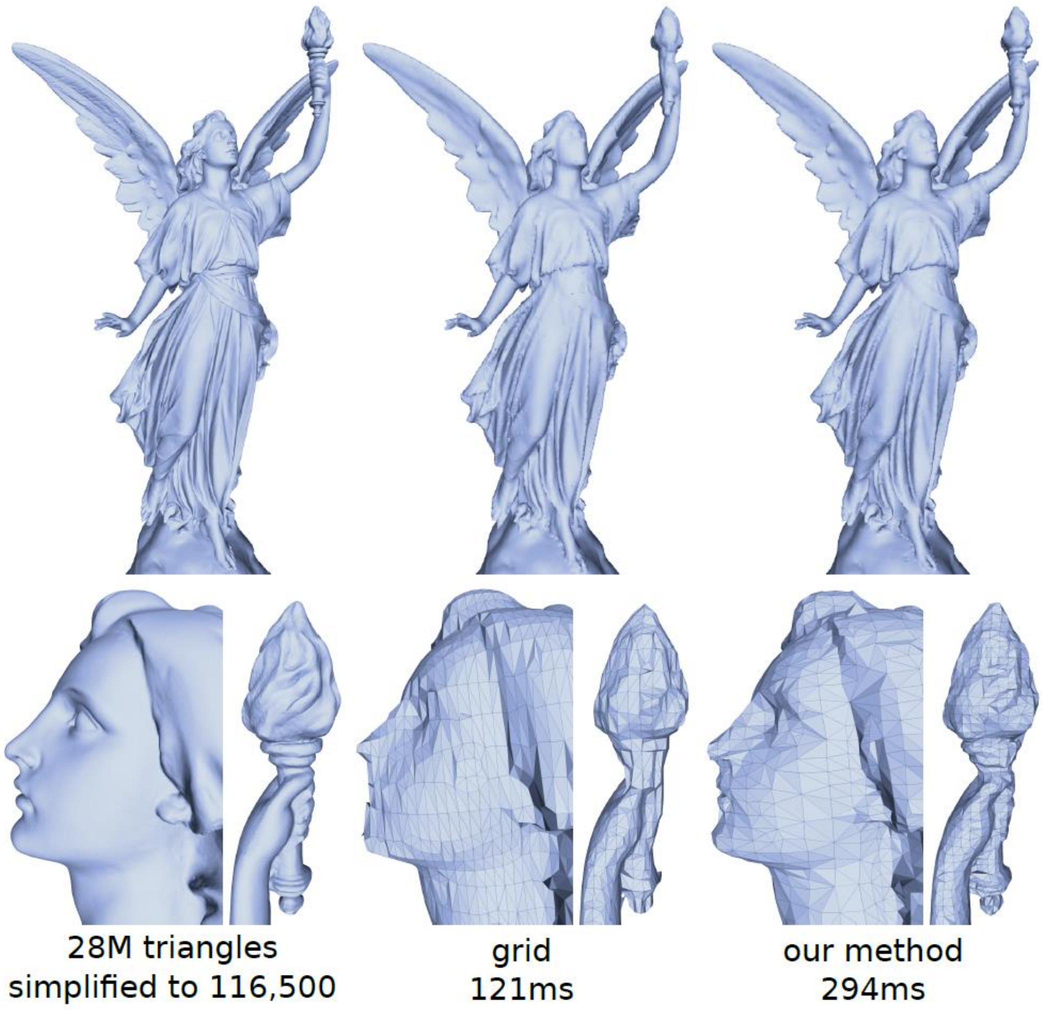

Figure 1: Instantaneous surface simplification for using WP3 F-SIMP program. The input model (left) is downsampled by several order of magnitude while keeping important geometric features in a quarter of a second using a single GPU.

The quest for high performance geometry processing went also toward higher levels operators, in particular the ones that help understanding the acquired data better. To this end, efficient geometry analysis methods have be developed to simplify the topology or the overall shape of an object for instance.

1.2.3 WP4: Integration of Modalities

Within the Harvest4D project, data coming from heterogeneous data sources and being of different modalities, such as range maps or color images, should be integrated into a consistent model of the world. During the first project year the efforts to achieve this key objective have concentrated on the foundational aspects of cross-modal integration (Task 4.1), particularly on statistical correlations between different modalities as well as representations that act as a bridge between different modalities. Based on the outcomes, initial research on algorithms for aligning data from multiple modalities (Task 4.2) and on cross-modal data compression (Task 4.3) has begun.

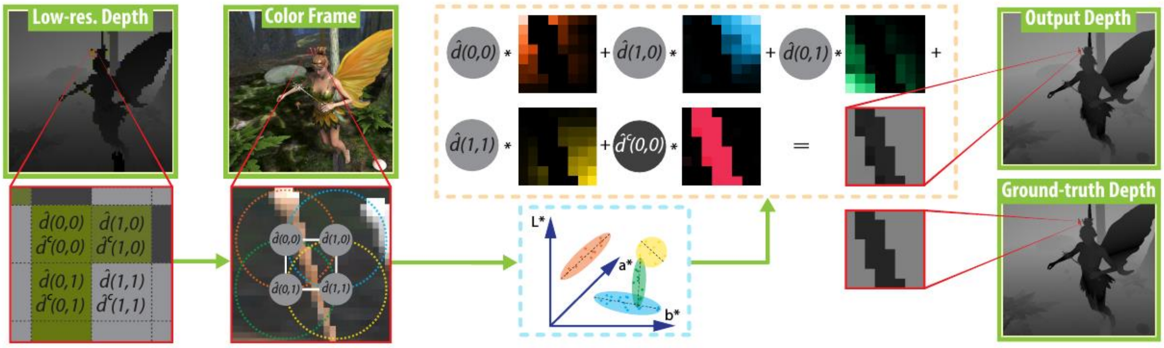

Figure 2: Basic pipeline from [Pajak et al. 2014] for reconstructing high resolution depth maps from low resolution depth maps using information from color images.

Concrete results include an analysis of the correlations between depth maps and color images, which is geared toward enabling novel depth compression algorithms. Figure 2 illustrates the approach, which can, for example, be applied to the widely used H.264/MPEG-4 AVC codec for video compression. Moreover, two complementary techniques for exploiting cross-modal dependencies between (color) images and 3D geometry have been developed, with initial applications in image-to-geometry alignment. In the future, they are aimed to serve as basis for cross-modal integration beyond images and geometry.

1.2.4 WP5: Multi-Scale Scene Processing

Work Package 5 revolves around the usage of scale at which individual samples are captured in order to derive novel algorithms for geometry processing. The per-sample scale value essentially describes the size of the surface regions from which the sample has been acquired. Further, real-world data is contaminated with noise of varying magnitude, which we would like to separate from the accurate geometric data.

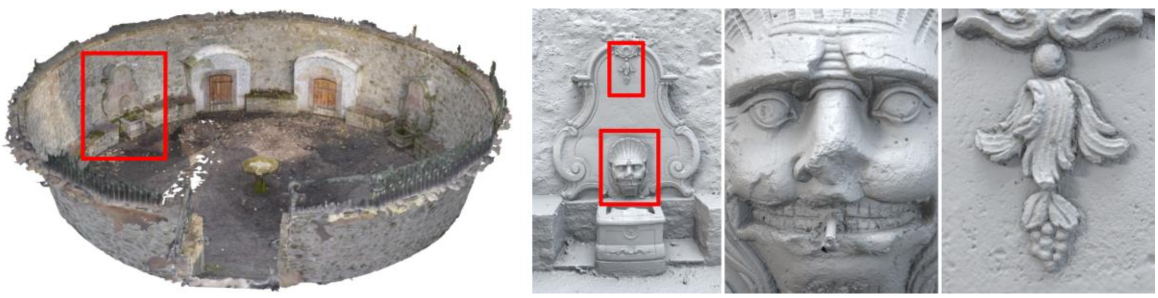

Figure 3: Multi-Scale scene reconstruction. The full, colored mesh (left) and some close-ups (right) with fine details.

The objective in Task 5.1 is to derive a theoretical foundation to use the scale values attached to the input samples (derived from the properties of the acquisition device) when combining multi-scale input samples into a single representation, such as a multi-resolution surface mesh. In particular, we explored the possibilities of frequency-space surface reconstruction techniques for merging surfaces at different scales into a combined representation. Work on Task 5.2 is concerned with practical approaches for multi-scale surface reconstruction (see Figure 3). In real-world scenarios the captured input data is contaminated with noise and outliers, which should ideally be separated from static and dynamic scene parts. In Task 5.4, we therefore started work on a 3-level decomposition technique to separate these individual components.

1.2.5 WP6: Detecting Changes

The main objective of this WP is the development of new algorithms that, starting from registered multimodal data (3D geometries from multiple sensors and images), are able to remove inter-device differences, to reconstruct a final dense geometry and to detect changes over time both in appearance and in geometry, by distinguishing in the acquired data the errors due to sensor inaccuracies as well as (reoccurring) changes in geometry, lighting, and environmental conditions.

Currently most of the work has been done in the Task 6.1 about detecting the changes in time varying 3D datasets (see Figure 4) and in Task 6.3 about dense geometric reconstruction. The main contribution in the Task 6.1 is a novel technique for the detection and the segmentation of the dynamic part of the input point clouds using implicit multi-scale surface descriptors, identifying and modeling the transient and permanent portions of a scene.

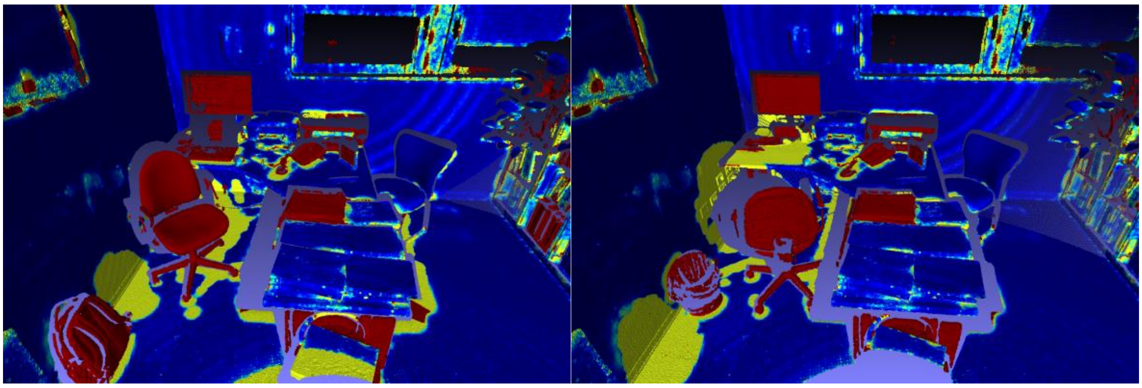

Figure 4: Color mapping of the change quality values in the geometry of two different acquisitions of the same environment in different time. The areas with a red color are the changes detected in the scene.

Works on Task 6.3 are concerned about the dense geometry reconstruction with a new multi-view technique for the reconstruction of mirroring objects and a new solution for dense 3D geometry and dense 3D motion reconstruction from stereo video streams.

1.2.6 WP7: Reflectance Reconstruction

The key objective of this WP is the development of novel techniques to separate illumination and reflectance. In particular, there is a need of taking into account shadows, reflections and camera characteristics to combine several measurements obtained under different illumination and capture conditions.

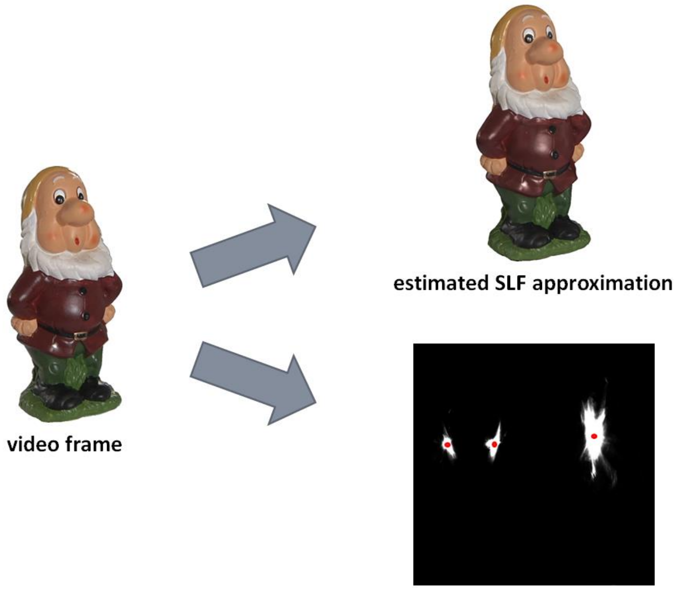

Figure 5: In the approach described in [Palma et al., 2013], both approximation of the surface light field and the main light directions are estimated from irregularly sampled data acquired with a moving camera.

To approach the task of separating illumination and reflectance, most of the work has been spent on material classification “in the wild”, surface light field estimation from sparse, irregularly sampled input data as obtained when using a moving video camera, and illumination estimation. The developed techniques are an important step towards the separation of illumination and reflectance in the incidentally data scenario where conditions are completely uncontrollable (see Figure 5).

1.2.7 WP8: Interaction and Visualization

Harvest4D has a strong focus on handling massive amounts of data. Consequently, these data sets are difficult to handle (Task 8.1), understand (Task 8.2), explore (Task 8.4), and query (Task 8.5). In particular, we are exploring image-based out-of-core rendering solutions (Task 8.1), which enable us to treat these otherwise unstructured data sets with a certain regularity. The task continues to the end of the second year. In the first part the work is mostly focused on static data sets whereas the second part will add the time component, a crucial element within Harvest4D. In parallel, visualization methodologies (Task 8.2) will be developed to offer a better understanding and interaction with time-dependent information. Ultimately, in the third year, we will explore specialized and modern hardware solutions for this purpose (Task 8.3). Due to the rapid changes in these areas, as well as the strong entanglement with Task 8.1 and 8.2, we already prepared first steps in this direction. One particular aspect is *remote rendering*, which will enable us to run our applications on almost any internet-enabled device, hereby supporting a widespread application of our results.

1.2.8 WP9: Integration and Evaluation

The objective of the WP is to ensure a progress towards an integrated prototype, demonstrating that the different blocks of work in the project contribute to the same proof-of-concept, avoiding fragmentation. The results obtained in the first years are:

- the design of a common standard for the development of code and the exchange of data
- the creation of some shared test datasets to be used for testing and evaluation

We expect to further expand the common datasets in order to provide more challenging examples. In the next year we will start the development of an integrated prototype based on some of the technologies developed in the project.

1.3 EXPECTED FINAL RESULTS

Overall, given the positive development of the project in the first reporting period, we expect the Harvest4D project to achieve its goal of making important first steps towards solving the problem of harvesting dynamic 3D data from incidental data capture. In particular, given the good progress on integration already in the first 12 months, delivering an integrated prototype that demonstrates how the individual components researched in the scientific work packages WP3-WP8 can work together seems well in reach. More concretely, we will in the following give more details for the expected final results in each scientific work package.

In **WP3**, we expect to contribute high performance processing, and analysis methods for dense 3D geometry, which can either come in the form of point clouds or polygonal meshes. The capability to provide high level analysis methods building on top of instantaneous geometric operators will be key for handling dynamic/evolving 3D data stream in a reasonable amount of time. After one year of research, interesting questions raised about the ideal underlying shape representation that could be the basis or the outcome of the work package. Beyond point cloud and meshes, higher level or more structured representation have already be addressed within the project, with a particular interest in medial structures (T3.1), Gaussian Mixtures Models (T3.1 and sparse voxel data structures.

In **WP4**, we expect to contribute accurate and efficient algorithms for registering and compressing data from heterogeneous data sources to the public that will on one hand enable obtaining time-varying 3D models of the world from incidental data alone, and on the other hand also enable other future applications where visual data from multiple sources must be processed together.

In **WP5**, we expect to contribute the methodology and algorithms to enable correct and mathematically well-founded processing of multi-resolution data captured from heterogeneous sources. This includes geometry, texture, as well as reflectance data and covers the complete pipeline from multi-resolution input to intermediate representations to final output.

In **WP6**, we expect that we will develop new algorithms that, starting from registered multimodal data, are able to detecting changes both in appearance and geometry over time and, eventually, to remove inter-device differences, allowing to reconstruct dense geometry in a more accurate way.

In **WP7**, we expect to contribute novel techniques to separate illumination and reflectance taking into account shadows, reflections and camera characteristics. The latter is necessary to combine several measurements obtained under different illumination and capture conditions which are expected in an incidentally captured data scenario.

1.4 PROJECT INFORMATION

All information on the Harvest4D project, including the work plan, regular updates on our progress, a list of publications, and examples of our work, can be found on the Harvest4D project page:

- http://www.harvest4d.org

The project is coordinated by Prof. Michael Wimmer from the Vienna University of Technology.
The partners are:

- Vienna University of Technology – Vienna, Austria (Lead: Michael Wimmer)
- Universität Bonn – Bonn, Germany (Lead: Reinhard Klein)
- Technische Universität Darmstadt – Darmstadt, Germany (Lead:Michael Goesele, Stefan Roth)
- Telecom ParisTech – Paris, France (Lead: Tamy Boubekeur)
- Consiglio Nazionale delle Ricerche (CNR) – Pisa, Italy (Lead: Paolo Cignoni)
- Technical University Delft – Delft, The Netherlands (Lead: Elmar Eisemann)

For any questions, contact

Project coordinator: Michael Wimmer
Phone: +43 1 58801 18687
Fax: +43158801 18698
E-mail: harvest4d-info@cg.tuwien.ac.at