HARVEST4D

HARVESTING DYNAMIC 3D WORLDS FROM COMMODITY SENSOR CLOUDS

Final integrated prototype evaluation

Deliverable 9.6

Date:	21.6.2016
Grant Agreement number:	EU 323567
Project acronym:	HARVEST4D
Project title:	Harvesting Dynamic 3D Worlds from Commodity Sensor Clouds

Document Information

Deliverable number	D9.6
Deliverable name	Final integrated prototype evaluation report
Version	1.0
Date	2016-06-21
WP Number	9
Lead Beneficiary	VUT
Nature	R
Dissemination level	PU
Status	Final
Author(s)	TUDA, VUT

Revision History

Rev.	Date	Author	Org.	Description
0.1	21/06/2016	S. Aroudj	TUDA	First draft
0.2	xx/xx/2016	M. Wimmer	VUT	Second draft
0.3	27/06/2016	All Partners		Integration of partners' individual prototypes
0.4	30/06/2016	M. Wimmer	VUT	Final version

Statement of originality

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

TABLE OF CONTENTS

1 EXECUTIVE SUMMARY

This document reports the evaluation of the integrated prototype for Work Package 9 (WP9), **Integration and Evaluation**, at the end of month 36 of the project Harvest4D.

The objective of this deliverable is to describe the evaluation of the final prototype defined in deliverable 9.5 (D9.5). As described in D9.5, the final prototype comprises a set of selected WP prototypes which are subdivided in two categories, prototypes involved in data processing and prototypes for visualization of intermediate or final data. There were three distinct meetings, so-called **Tech Parties,** to gather data for the testing and evaluation of the mentioned prototypes and facilitate partner collaboration as well as knowledge exchange.

The following sections first present what and how data for the evaluation was gathered. Then they compactly present the evaluation results for each prototype ordered by category and work package.

2 EVALUATION DATA

2.1 THE ARENE DE LUTECE DATASET

In each year of the project, there was a special meeting called **Tech Party**. Each of these three meetings included a capture session to gather uncontrolled (i.e., "incidental") data of the site *Arène de Lutèce* in a large-scale Harvest4D fashion. The French site is shown in Figure 1. Besides research presentations for collaborative prototype understanding, the meetings also included discussions to define reasonable data file formats and prototype interfaces to enable the data flow of the final integrated prototype of D9.5.

During every Tech Party, the participants were sent out for capturing of the target site in a "Harvest4D fashion". That means there was no acquisition plan. Attendees mainly took photos like tourists would do it. That is why the three capture sessions resulted in a very challenging data set.

Figure 1. Example image of the site *Arènes de Lutèce,* which was captured during each Tech Party.

First of all, the amount of captured data is huge, since several thousands of unregistered images of the French arena were taken during each meeting. The participants took 13,169 photos of *Arènes de Lutèce* in total.

Second, the image quality varies strongly. The participants mainly used low-quality sensors, such as mobile phone cameras. In contrast to that, a few attendees used high-end capture devices. Regardless of the capture device, a subset of the taken photos exhibits artifacts of varying severity like motion blur, wrong depth of focus, over- or underexposure, lens distortion effects etc.

Third, there are areas which were captured intensively, so that there are a lot of redundant data points for some parts of the data set. In contrast, there are also less interesting spaces for which only little data was gathered.

Fourth, there are large differences in surface sample scales. On the one hand, boring surfaces like the arena ground were mostly captured from large distances resulting, in relatively inaccurate and coarse surface samples. On the other hand, interesting parts like a statue or an ornamented water dispenser were digitized by close up photos.

Fifth, there are multi-type as well as multi-scale scene changes between the captured images. With respect to the type of change, illumination conditions, scene geometry or surface materials exhibit variations over time. Regarding scale, there are small-scale changes due to the fact that a single capture session was done during multiple hours of a single day. This caused for example different illumination conditions in the pictures created by sun movement. Apart from that, there was always a time gap of roughly a year between successive acquisition sessions to enable capturing of large-scale changes of the arena scene, such as moved objects, weathering effects, different vegetation or distinct weather conditions.

The captured data, intermediate and final results are publically available on the Harvest4D web page under the menu “Data Sets”.

2.2 STATISTICS OF THE DATA SET

Statistics	2014	2015	2016
Number/size of photos	1,456 (10.8 GB)	6,052 (15 GB)	5,663 (30.4 GB)
Number of people/sensors	6 (5)	10 (9)	8 (8)
# Points/size in dense point cloud	60.5M (1.7 GB)	301MM (8.5 GB)	29.1 GB (in 3 LODs)
#Vertices/Size of FSSR-generated reconstruction	3.6M (57 MB)	12.8M (686 MB)	16.5M (2 GB)

3 EVALUATION OF DATA PROCESSING PROTOTYPES

For better understanding, please recall the final integrated prototype and its data flow, presented in Figure 2. Green boxes represent data, while blue ones depict individual WP prototypes. In the following sub sections, the evaluation of each individual prototype is discussed.

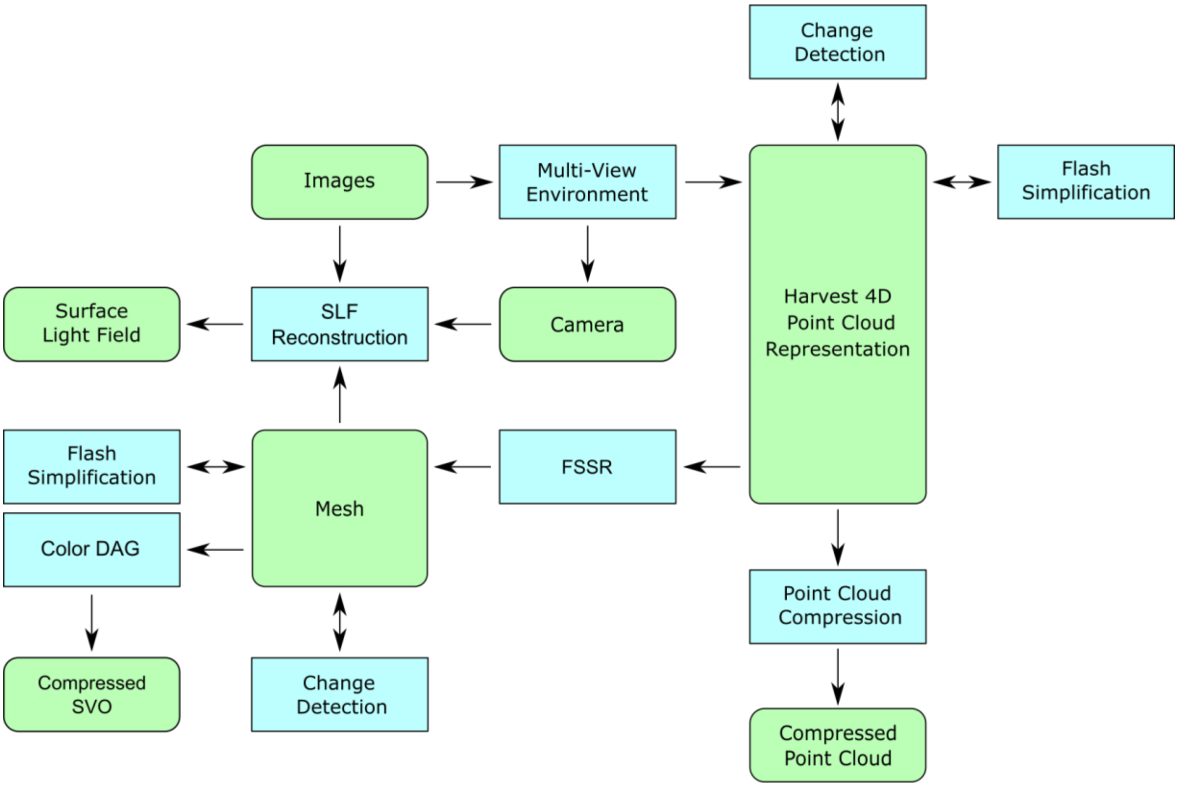

Figure 2. Overview of the final integrated prototypes regarding data processing.

3.1 MULTI-VIEW ENVIRONMENT

The prototype called Multi-View Environment (MVE) [Fuhrmann et al. 2014a], described in D5.31 and D5.32, was successfully used to reconstruct undistorted images and depth maps. Figure 3 shows an undistorted example input image and its associated depth map taken from the input set of the second Tech Party, which comprises 6051 pictures.

Figure 3. Undistorted image (left) and reconstructed depth map (right) of an ornamented fountain at the arena site.

Furthermore, we reconstructed sparse and dense 3D point clouds of the complete scene as well as camera distributions for each capture session. Figure 4 presents the sparse point cloud reconstruction and calculated camera distribution of the meeting in the year 2015.

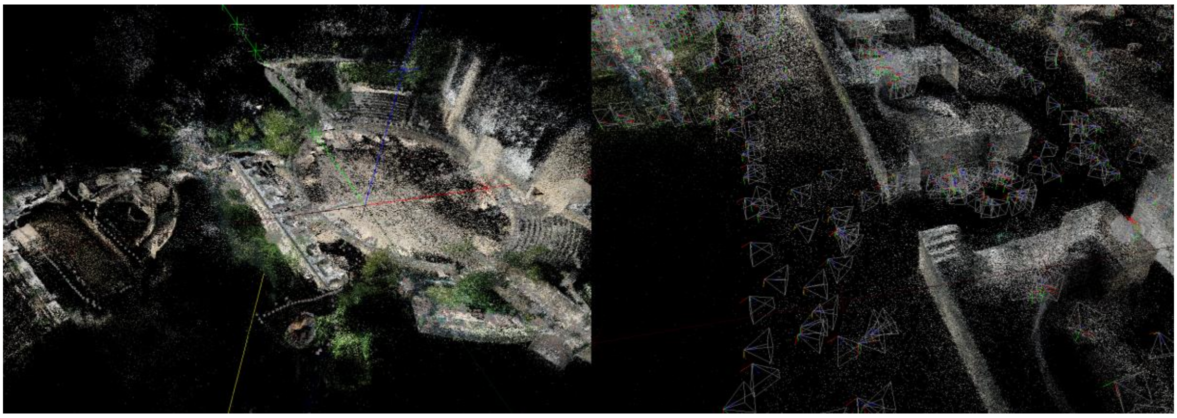

Figure 4. Zoomed out (left) and closer view (right) of the sparse point cloud and camera poses for the data of 2015.

The reconstructions show that the MVE prototype can robustly handle the uncontrolled and challenging data set of the arena. MVE was used to calculate large, complex and globally consistent multi-view models, which in turn were processed by many other individual prototypes.

3.2 FLASH SIMPLIFICATION

The prototype called HSGS, which implements Morton Integrals for high speed geometry simplification [Legrand and Boubekeur 2015], was successfully used to reduce the size of the objects in the Paris Arene de Lutèce dataset. This automatic simplification process was able to account for the color values captured by MVE and provide simplification at different levels, reducing by one to two orders of magnitude the sample count while maintaining the main geometric structures of the scene.

Figure 5 Simplification in real time from the raw model (left) to 10, 5 and 3% of its original resolution with HSGS

HSGS was experimented both in mesh or point versions, and ran on both multicore CPUs and GPUs. Each time, shapes such as the central statue in the Arene de Lutece were obtained at a desired level of detail instantaneously, without any precomputation, directly from the raw MVE reconstruction (see Figure 5). More concretely, on an NVIDIA Geforce 980TI, the statue shown above, consisting of 1,529,256 input polygons, was decimated to 10% in 14 ms. Decimations to 5% and 3% take equally long. When run on the full Arena scene with 7,111,435 input polygons, decimation to 10% takes 36 ms (see Figure 6).

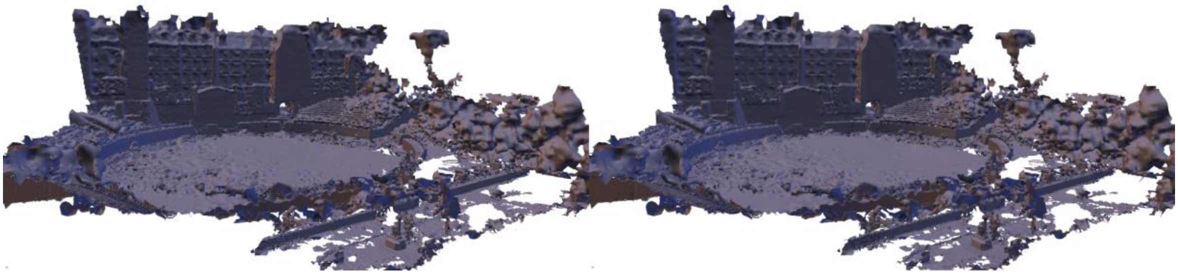

Figure 5 Simplification off the full Arena scene in real time from the raw model (left) to 10% of its original resolution with HSGS

3.3 FSSR (FLOATING SCALE SURFACE RECONSTRUCTION)

The FSSR prototype [Fuhrmann et al. 2014b], which is described in D5.31 and D5.32, helped to successfully accomplish the reconstruction of surface meshes for the large site *Arènes de Lutèce.* Figure 7 shows that FSSR supports creation of a globally consistent surface mesh for an input point cloud consisting of surface samples with strongly varying scale.

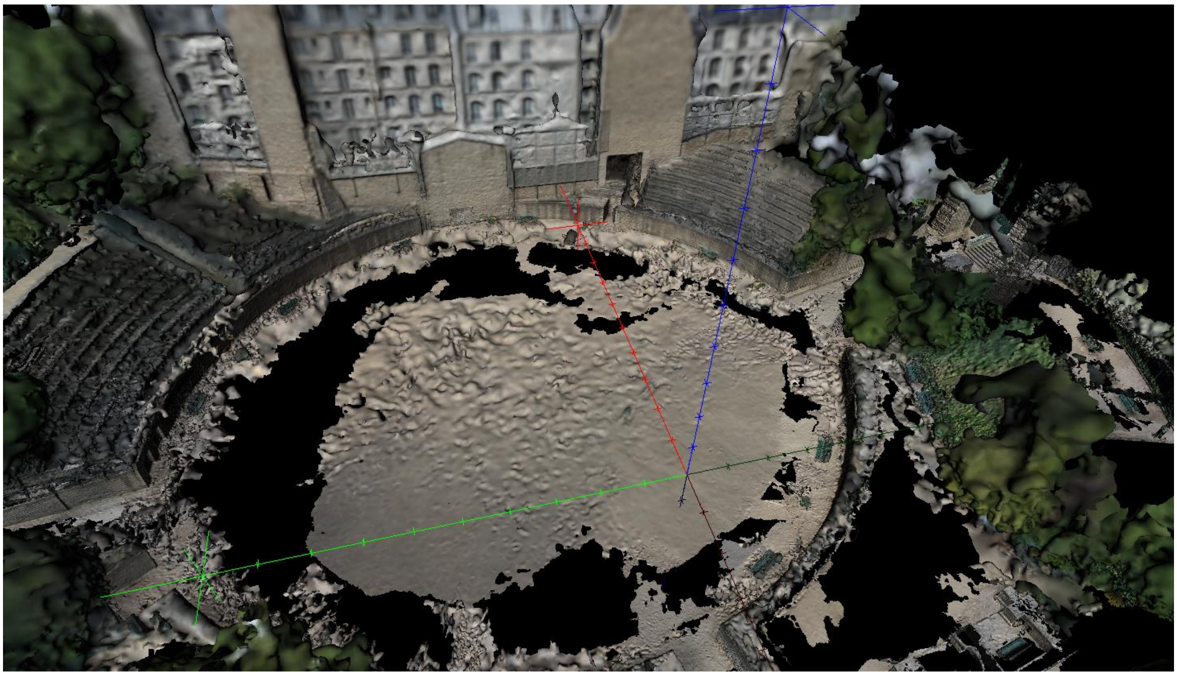

Figure 7. Multi-scale surface mesh of the capture session in 2015, which was produced by means of FSSR.

Figure 8 demonstrates that FSSR is able to cope with strong differences in surface sample scale by presenting very coarse and extremely fine levels of detail of a wall that was quite interesting for a capture session attendee and therefore captured from close-up.

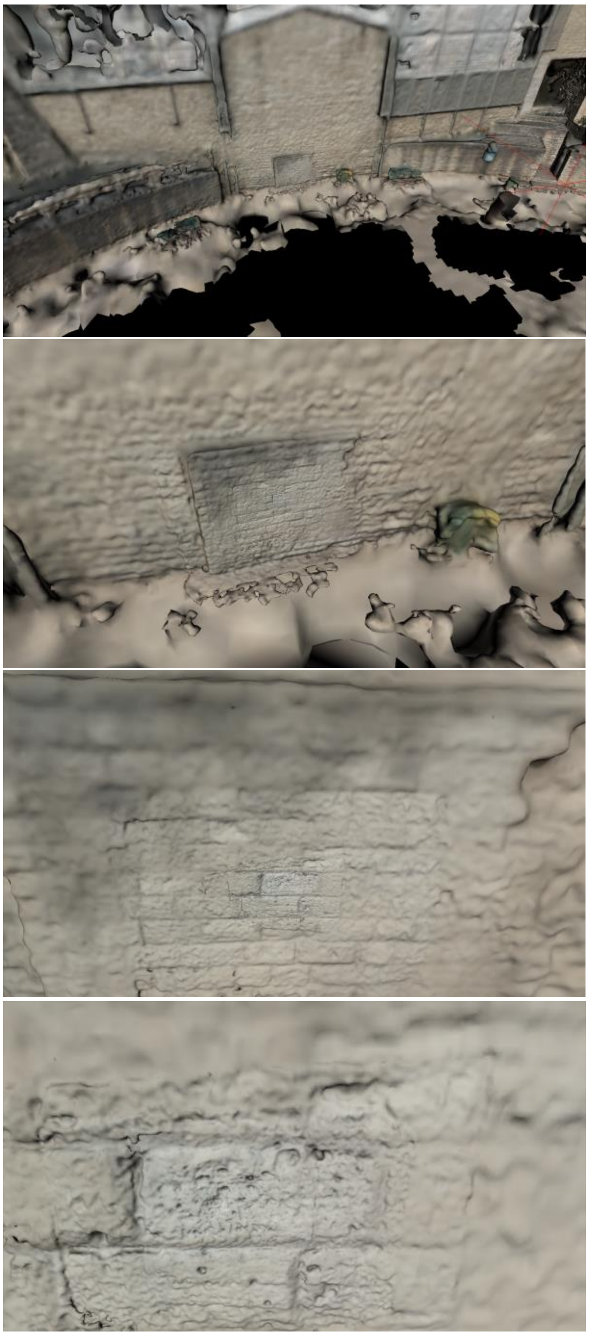

Figure 8. Increasing level of detail of the multi-scale reconstruction of a house wall at the arena site.

3.4 CHANGE DETECTION

The change-detection prototype [Palma et al. 2016], which is described in D6.11 and D6.12, was used to compare a pair of 3D reconstructions of the *Arènes de Lutèce* site computed using the MVE and the FSSR prototypes on the datasets captured in two different Tech parties. The prototype computes a continuous per-vertex change field that can be used to segment the input

3D models between the static and the dynamic regions of the scene. In order to be more robust to noise, we apply the prototype on the 3D meshes computed with the FSSR prototype instead of the raw point cloud. Figure 9 shows the binary segmentation of the two models (blue = no-change, red = change).

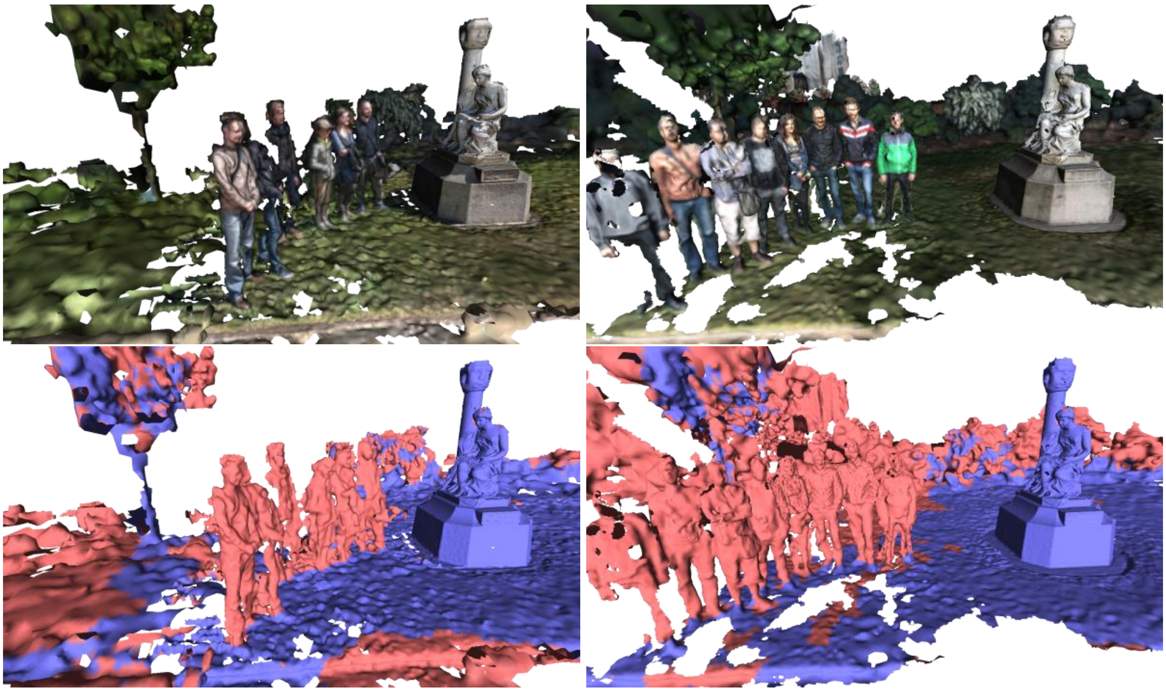

Figure 9. (Top) Rendering of the 3D mesh of two different captures of the Arene de Lutece site. (Bottom) Binary segmentation of the changed regions between the two captures.

The prototype was tested to compare the triangular meshes produced using the data of the three capture sessions of the Arene de Lutece, composed by 3.5 Milion (capture 2014), 12 Milion (capture 2015) and 16 Milion (capture 2016) vertices. In particular, we compare the 2014 capture with the 2015 and the 2015 capture with the 2016. The amount of geometry with a probability higher than 50% to be a change is about the 45% for the pair 2014-2015 and about 53% for the pair 2015-2016. The computation takes about 2 minutes for the pair 2014-2015 and about 3 minutes for the pair 2015-2016 to compute the per-vertex continuous change field. We use only the first part of the algorithm described in [Palma et al. 2016] due to the high level of noise around the borders of each mesh, which prevents the consolidation of the change information during the propagation step.

3.5 POINT CLOUD COMPRESSION

The point cloud compression prototype [Golla & Klein, 2015], described in D4.31 and D4.32, was used to obtain a compressed representation of colored point cloud data of the *Arènes de Lutèce* site obtained using the MVE and FSSR prototypes on the datasets captured in two different Tech parties. The compressed data computed by the prototype consists of an occupancy map, a height map and the respective color information for local patches into which the surface is decomposed (see Figure 10). Decompressing this compressed data yields again a point cloud representation (see Figure 11). For the shown example, the compression ratio is about 1:55. Such compression tasks are particularly important when many scenarios have to be archived.

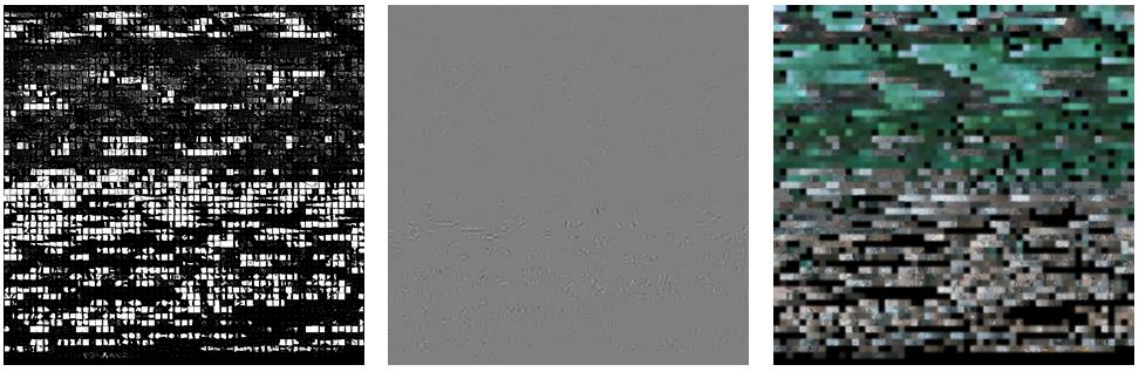

Figure 10. (Left) Occupancy map, (middle) height map and (right) color map for the per-patch information.

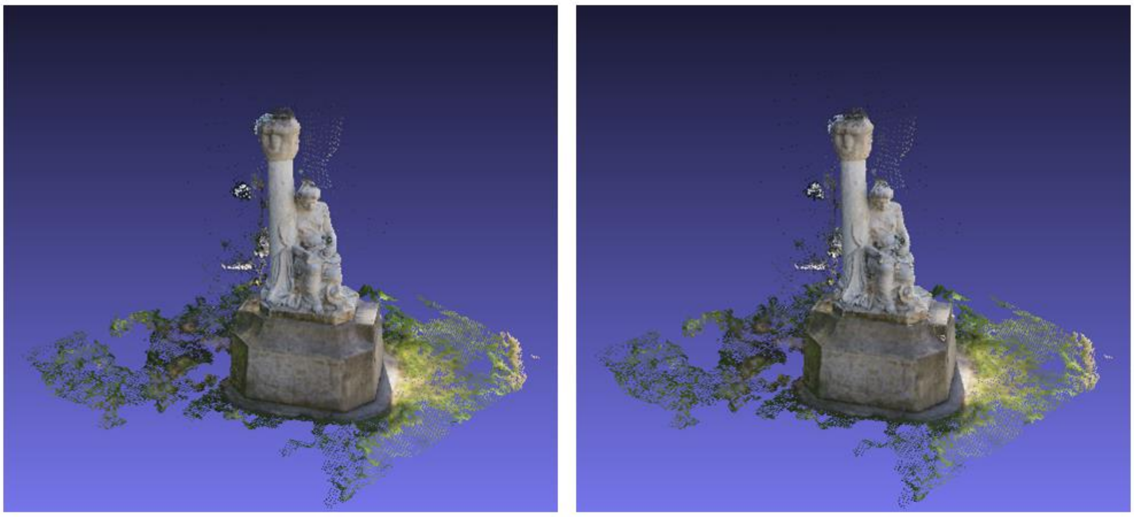

Figure 11. (Left) Original point cloud representation and (right) point cloud obtained after decompressing the compressed data.

3.6 COLOR DAG

The color DAG prototype [Dado et al. 2016] is described in D4.41 and D4.42. It enables a very high-resolution voxel representation of the *Arènes de Lutèce* scene in a hierarchical SVO structure that is subsequently compressed by an order of magnitude. Thanks to the compression, we can store the scene fully in-core, which easily enables real-time rendering.

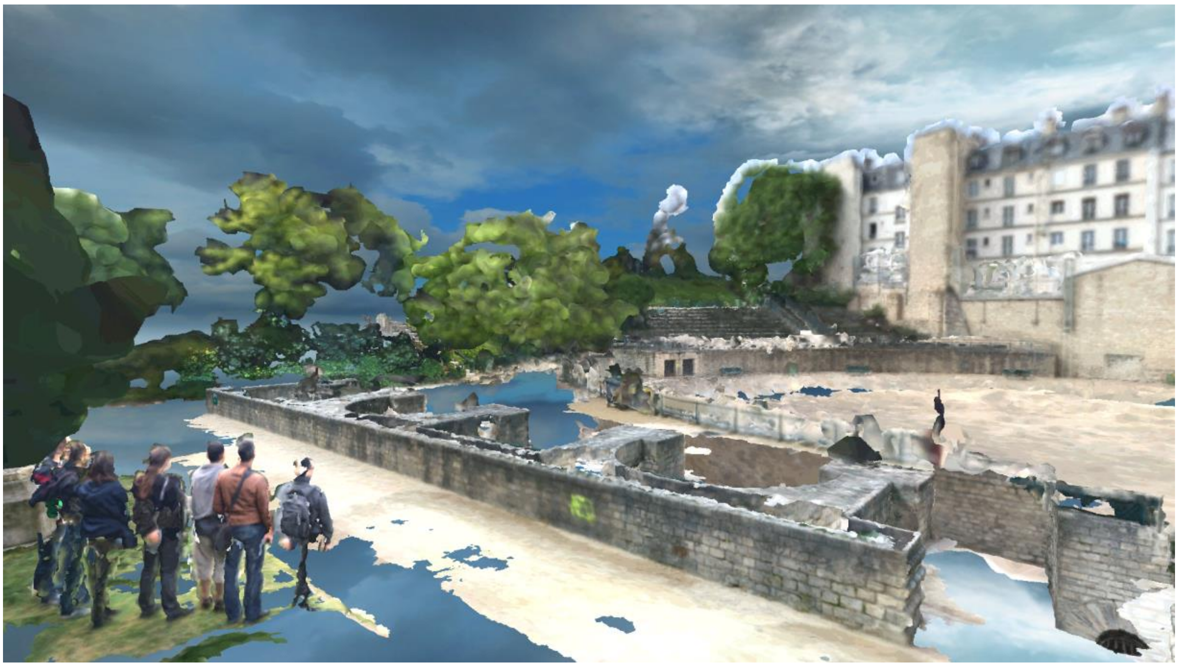

Figure 12. High-resolution voxel representation of the *Arènes de Lutèce* scene. We are able to store the scene in core at a 128K³ resolution.

Figure 12 shows a snapshot of our high-resolution voxel representation. The small size of the voxels hides their structure, since generally, a voxel is projected to an area smaller than a pixel.

The color DAG can be applied on any OBJ or PLY file, which means in our integrated prototype, it is used on the meshes generated by FSSR.

Our prototype is able to compress the voxelized 2015 Arena scene at a $64K^3$ resolution down to a size of 1 GB, while a complete SVO would require nearly 10 GB. The original triangle mesh contained 12 million vertices. The compression algorithm itself is however highly unoptimized and currently takes a day for scenes of this magnitude. Still, we enable real-time rendering performance with a framerate around 30FPS.

3.7 SLF RECONSTRUCTION

The SLF reconstruction prototype [Palma et al. 2013], described in D7.21 and D7.22, was used to estimate the Surface Light Field of a 3D mesh to approximate its reflectance behavior. More specifically, the Surface Light Field allows the rendering of the appearance of the object from different viewpoints but with the same lighting conditions, i.e., the same conditions used during the acquisition of the input photos. Since we need several color samples per each point of the surface, we select a subset of the photos of the capture sessions of the *Arènes de Lutèce* site related to objects acquired with a high number of images, namely a statue and a small fountain. For each dataset, the prototype takes the 3d mesh computed with the MVE and FSSR prototypes and the camera parameters of each photo, exported by the MVE prototype, to project the color data on the 3D surface. The estimated Surface Light Field is composed by two terms: the diffuse color and the other reflectance residual effects. The processing time was about 5 minute for the biggest dataset composed by 393 photos and a mesh of 2.5 million of vertices. The rendering of the computed surface light field is real-time using a recent PC. Figures 13 and 14 show the results for the two tested datasets.

Figure 13. (Left) Rendering of the Surface Light Field of the fountain. (Center) Diffuse component. (Right) Residual component.

Figure 14. (Left) Rendering of the Surface Light Field of the statue. (Center) Diffuse component. (Right) Residual component.

4 EVALUATION OF VISUALIZATION PROTOTYPES

This section discusses prototypes which are solely or partially dedicated to output data visualization and which are thus related to Work Package 8. For better understanding, Figure 15, taken from D9.5, shows an overview of the associated input data that these visualization prototypes can handle.

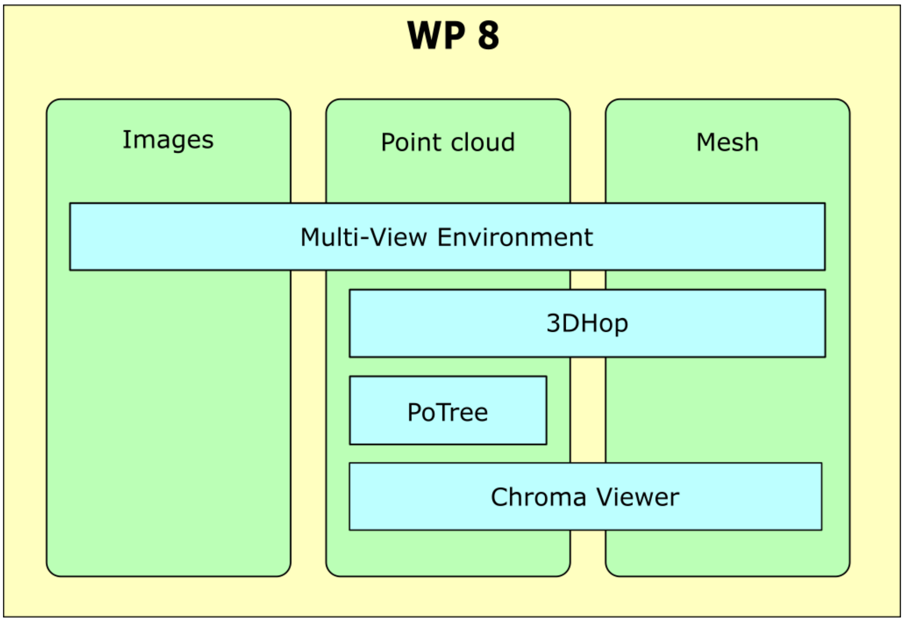

Figure 15. Overview of visualization prototypes of WP 8 and their associated data.

4.1 MULTI-VIEW ENVIRONMENT

The graphical user interface of MVE, named Ultimate Multi-View Environment (UMVE) [Fuhrmann et al. 2014a], is able to visualize a variety of data sets of different type in a user-friendly manner. For example, it allows the user to easily inspect Structure from Motion (SfM) data, like sparse point clouds with associated cameras as depicted by Figure 16. Besides undistorted images, depth maps and other view-dependent data, it is able to render late stage intermediate results like surface triangle meshes calculated by FSSR.

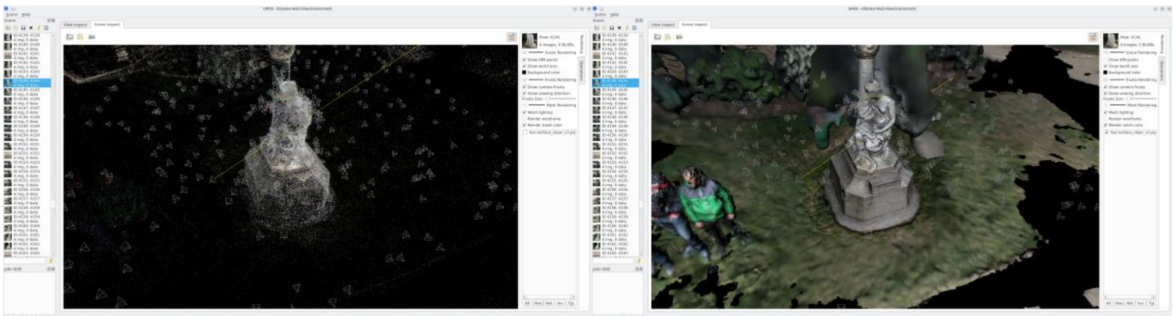

Figure 16. Visualization of SfM results (left) and the arena surface mesh with cameras (right) from year 2015.

Thanks to the user-friendly visualization features of UMVE, it was possible to quickly examine the complex arena scene data and detect problems regarding different individual reconstruction steps. This fertilized research discussions and work improvements.

4.2 3DHOP

The 3DHOP prototype [Ponchio et al. 2015] allows the publishing on the web of a huge 3D model using a compressed multi-resolution 3d format that optimizes the transmission of the data needed for a view-dependent rendering of scene. The multiresolution format and the compression scheme are described in D8.31 and D8.32. The prototype is used to develop a new graphics interface that allows the space-time navigation of a scene with the purpose to highlight the most significant changes. Starting from the output computed with the change-detection prototype, we computed the multi-resolution format of each model by mapping the change fields to the alpha channel. This allows creating a web page for the interactive exploration and navigation of the two models. Figure 17 shows the graphical user interface for the space-time navigation of the scene using a simple slider to explore the temporal changes.

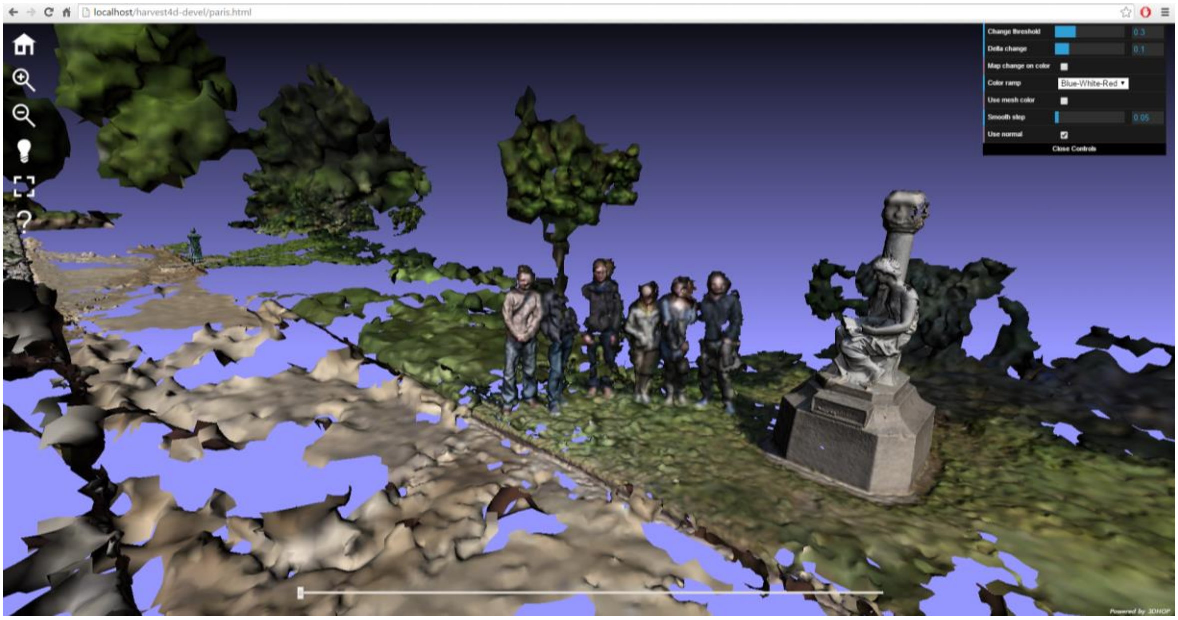

Figure 17. Web page with the space-time navigation interface to explore two different captures of Arene de Lutece.

4.3 POTREE

Potree is a remote visualization tool that runs only in a web browser (Figure 18). It focuses purely on point clouds (meshes are not supported) and runs on all recent browsers supporting WebGL. Potree is focused on rendering very large point clouds. It uses an internal hierarchical octree representation which is computed in a preprocessing step. Compared to 3D HOP, it is focused on different visualization modes, including screen-aligned splats, geometric splats, and Gaussian splats, to explore different quality/performance tradeoffs. Potree also allows handling arbitrarily large point clouds.

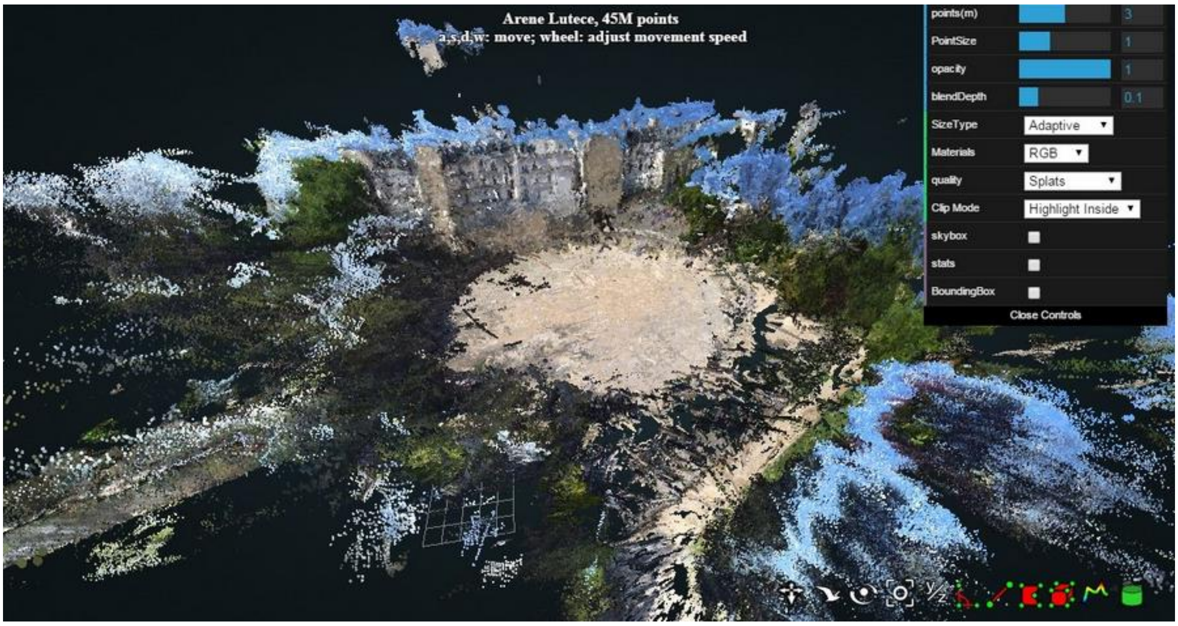

Figure 18: The Potree viewer uses an adaptive point size for best quality.

4.4 CHROMA VIEWER

The chroma viewer [Schemali and Eisemann 2014] is described in D8.21 and D8.22. It enables a better depth perception by simply coloring a scene based on a surface's distance to the camera. Nearby objects are colored red, far away objects blue, and objects in between get a white color. We linearly interpolate the colors for a smooth transition. When viewing such colored scenes with the ChromaDepth glasses, they appear in 3D, while still enabling exploration without the glasses as well. For the arena scene, this means we can better understand the topology of the scene. The coloring is straightforward and extremely efficient to implement on the GPU, making the performance overhead near negligible (see Figure 19).

Figure 19: The chroma viewer provides improved depth perception when wearing the ChromaDepth glasses.

5 SUMMARY

This document has shown that all components of the final integrated prototype were successfully carried out on a challenging data set. Arene de Lutece was captured under exactly those conditions that we initially set out to tackle in this project: incidental capture using different devices at different times. The resulting data set is very large, but can still be processed quickly using some of our tools due to the advancements of WP3 (e.g., flash simplification). The data of different quality and different times was registered using MVE (WP4), and a high-quality reconstruction took multiple scales into account (WP5). Different web viewers show the results (WP8), even including a visualization of the detected changes (WP6). Another module allows a high-quality depiction of material properties if enough samples are available through surface light fields (WP7).

This data set has therefore allowed us to demonstrate results for all work packages of this project. While a detailed numerical evaluation is beyond the scope of this document, each component is based on a scientific publication where such an evaluation is available is available.

Furthermore, we have decided to make this data set publically available so that other researcher can benefit from our efforts. This will allow them to test their own reconstruction approaches, and compare with the reconstructions that we produced using our pipeline, which we also make available. The prototype D9.5 itself is also publically available, so the results can be reproduced and compared to other approaches.

6 REFERENCES

- [Dado et al. 2016]
  Bas Dado, Timothy R. Kol, Pablo Bauszat, Jean-Marc Thiery and Elmar Eisemann
  Geometry and Material Compression in High-resolution Voxelized Scenes
  In: Eurographics, 2016

- [Fuhrmann et al. 2014a]
  Simon Fuhrmann, Fabian Langguth and Michael Goesele
  MVE – A Multi-View Reconstruction Environment
  In: Proceedings of the Eurographics Workshop on Graphics and Cultural Heritage, 2014.

- [Fuhrmann et al. 2014b]
  Simon Fuhrmann and Michael Goesele
  Floating Scale Surface Reconstruction
  In: ACM Transactions on Graphics (Proceedings of SIGGRAPH), 2014.

- [Golla et al. 2015]
  Tim Golla and Reinhard Klein
  Real-time Point Cloud Compression
  In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2015

- [Legrand et al. 2015]
  Hélène Legrand and Tamy Boubekeur
  Morton Integrals for High Speed Geometry Simplification
  In: High Performance Graphics, 2015

- [Palma et al. 2013]
  Gianpaolo Palma, Nicola Desogus, Paolo Cignoni and Roberto Scopigno
  Surface Light Field from Video Acquired in Uncontrolled Settings
  In: International Conference DigitalHeritage, 2013

- [Palma et al. 2016]
  Gianpaolo Palma, Paolo Cignoni, Tamy Boubekeur and Roberto Scopigno
  Detection of Geometric Temporal Changes in Point Clouds
  In: Computer Graphics Forum (In press), 2016

- [Ponchio et al. 2015]
  Federico Ponchio and Matteo Dellepiane
  Fast decompression for web-based view-dependent 3D rendering
  In: International Conference on 3D Web Technology, 2015

- [Schemali et al. 2014]
  Leila Schemali and Elmar Eisemann
  ChromoStereoscopic Rendering for Trichromatic Displays
  In International Symposium on Non-Photorealistic Animation and Rendering, 2014.

- [Schuetz et al. 2015]
  Markus Schuetz and Michael Wimmer
  High-Quality Point Based Rendering Using Fast Single Pass Interpolation
  In: International Conference DigitalHeritage, 2015