

This project is funded by the Horizon 2020 Framework Program of the European Union

Built2Spec

Built to Specifications – Tools for the 21st Century Construction Site

H2020 Grant Agreement – 637221

D4.3 – 3D Capture Deliverable

Primary Author: ETH

Contributors: R2M, EURECAT, TNO, NUIG

1st Quality reviewer: Myriam Servières (ECN)

2nd Quality reviewer: Daniele Bortoluzzi (R2M)

3rd Quality reviewer: Sata Casciati (R2M)

Deliverable nature:	Report (R)
Dissemination level (Confidentiality)	Confidential (CO)
Contractual delivery date:	M30 – June 30 th , 2017
Actual delivery date:	30 June 2017
Version:	5
Total number of pages:	45
Keywords:	3D Reconstruction, SfM, SLAM, Point Cloud, Mesh

DISCLAIMER

The opinion stated in this report reflects the opinion of the authors and not necessarily the opinion of the European Commission. Neither the EACI nor the European Commission are responsible for any use that may be made of the information contained therein.

All intellectual property rights are owned by the BUILT2SPEC consortium members and are protected by the applicable laws. Unless otherwise specified, all document contents are '@BUILT2SPEC – All rights reserved'. Reproduction is not authorized without prior written agreement.

The commercial use of any information contained in this document may require a license from the owner of that information.

All BUILT2SPEC consortium members are committed to publish accurate and up to date information and take the greatest care to do so. However, neither the European Commission nor the BUILT2SPEC consortium members can accept liability for any inaccuracies or omissions nor do they accept liability for any direct, indirect, special, consequential or other losses or damages of any kind arising from the use of this information.

ACKNOWLEDGEMENT

This document is a deliverable of the BUILT2SPEC project which has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement no. 637221.

ABSTRACT

The acquisition of accurate and reliable 3D data is an essential step for the major goals of the Build2Spec project as it is required for novel inspection and progress tracking methods, as well as for quality check and management tools. This deliverable gives an overview of various acquisition methods and 3D scanning technologies that have been applied within the project. In particular, active depth sensing technologies like laser scanners, Google Tango tablet and Microsoft HoloLens devices have been used. Furthermore, multiple passive depth sensing technologies in which 3D models are computed from large set of images which were captured by a human with a GoPro camera or images from flying drones have been processed.

Moreover, we discuss automatic building information modelling (BIM) from such 3D reconstructions as well as the integration of captured data into the virtual construction management platform (VCMP).

TABLE OF CONTENTS

ABSTRACT	3
TABLE OF CONTENTS	4
TABLE OF FIGURES	6
ABBREVIATIONS	7
1. INTRODUCTION	8
1.1 OVERVIEW	8
2. BACKGROUND	9
2.1 DEFINITIONS	10
2.1.1 <i>Point cloud</i>	10
2.1.2 <i>Mesh</i>	10
2.1.3 <i>BIM</i>	10
2.2 STRUCTURE FROM MOTION (SFM)	10
2.2.1 <i>Feature Extraction</i>	11
2.2.2 <i>Matching</i>	11
2.2.3 <i>Geometric Verification</i>	12
2.2.4 <i>Structure and motion reconstruction</i>	12
2.3 SIMULTANEOUS LOCALIZATION AND MAPPING (SLAM)	12
3. TECHNOLOGY DESCRIPTION	13
3.1 DEVICES DESCRIPTION	13
3.1.1 <i>GoPro</i>	13
3.1.2 <i>Drone (EURECAT & R2M)</i>	14
3.1.3 <i>Google Tango</i>	17
3.1.4 <i>Microsoft HoloLens</i>	18
3.1.5 <i>Leica Laser Scanner</i>	19
3.1.6 <i>Comparison of Devices</i>	19
3.2 3D RECONSTRUCTION SOFTWARE DESCRIPTION	21
3.2.1 <i>COLMAP (ETH)</i>	21
3.2.2 <i>Agisoft Photoscan</i>	21
3.2.3 <i>Pix4D</i>	23
3.2.4 <i>Comparison of 3D Reconstruction Software</i>	24
4. REAL DATA ACQUISITION	25
4.1 ITALIAN DEMO SITE "SAN SALVO" DATASET (R2M & DE5)	25
4.1.1 <i>Indoors</i>	25
4.1.2 <i>Outdoors</i>	27
4.1.3 <i>The results achieved from the different devices employed are in line with the expected results.</i> <i>Multiple rooms</i>	27
4.1.4 <i>Progress Tracking</i>	29

4.2	DATASET SCHEPENDOMLAAN (TNO).....	29
4.3	LASER DEMONSTRATOR (OPC & NUIG).....	31
4.4	EURECAT DATASET.....	32
4.5	ACQUA & SAPONE DATASET (R2M & DE5).....	32
5.	INTEGRATION WITH VCMP AND BIMSERVER	35
5.1	INTEGRATION WITH VCMP	36
5.1.1	<i>Inputs and Outputs</i>	37
5.2	EXAMPLE WITH BIMSERVER	38
6.	ADDITIONAL WORK	40
	RAPID BIM MODELLING FROM 3D RECONSTRUCTIONS (ETH).....	40
7.	CONCLUSIONS.....	42
	REFERENCES TO B2S DELIVERABLES:	43
	PUBLICATIONS SUPPORTED BY BUILT2SPEC PROJECT.....	44
	LITERATURE REFERENCES	45

TABLE OF FIGURES

FIGURE 1-1 OVERVIEW OF 3D DATA ACQUISITION AND DATA PROCESSING. DELIVERABLE D4.3 COMPRISES THE DATA ACQUISITION WITH VARIOUS TECHNOLOGIES. DELIVERABLE D4.4 DEALS WITH PROCESSING AND ANALYSING THE 3D DATA. 9

FIGURE 2-1 COLISEUM 3D RECONSTRUCTION, FROM FLICKR PHOTO COLLECTION. 9

FIGURE 2-2 MODEL REPRESENTATIONS ARE SHOWN: POINT CLOUD (LEFT), TRIANGLE MESH (MIDDLE), AND BIM (RIGHT). 10

FIGURE 2-3 OVERVIEW OF AN INCREMENTAL STRUCTURE-FROM-MOTION PIPELINE. 11

FIGURE 3-1 GoPRO HERO. 13

FIGURE 3-2 PATH PLANS FOR DIFFERENT CASES. FROM LEFT TO RIGHT. BUILDING RECONSTRUCTION AND CORRIDOR MAPPING. 15

FIGURE 3-3 OVERLAP BETWEEN DIFFERENT IMAGE ACQUISITION PLANS, WITH A RIGHT CONFIGURATION ON THE LEFT AND AN INCORRECT ONE ON THE RIGHT. 16

FIGURE 3-4 IMAGE ACQUISITION PLAN DESIGNED BY EURECAT FOR THE RECONSTRUCTION OF THE VIADUCT. A GRID PATTERN WAS USED FOR THE FIRST FLIGHT (RED LINE), AND TWO FLIGHTS WERE DONE AT THE FRONTAL AND REAR SIDE (GREEN LINES)..... 17

FIGURE 3-5 3D RECONSTRUCTION OF THE VIADUCT DONE BY EURECAT (PIX4D, LEFT) AND BY ETH (COLMAP, RIGHT)..... 17

FIGURE 3-6 GOOGLE TANGO TABLET. 18

FIGURE 3-7 GOOGLE TANGO APPLICATION: TANGO MAPPER..... 18

FIGURE 3-8 MICROSOFT HOLOLENS. 19

FIGURE 3-9 LEICA LASER SCANNER. 19

FIGURE 3-10 COLMAP'S INTERFACE..... 21

FIGURE 3-11 OVERLAPPING AND VIEWPOINTS (FROM AGISOFT PHOTOSCAN GUIDELINE). 22

FIGURE 3-12 PIX4D MAPPER INTERFACE..... 24

FIGURE 4-1 SAN SALVO - ITALIAN DEMO SITE. BIM (LEFT) AND NORTH-EAST VIEW (RIGHT). 25

FIGURE 4-2 GoPRO + COLMAP: 3D RECONSTRUCTION FROM GoPRO IMAGES AFTER PROCESSING WITH COLMAP 25

FIGURE 4-3 GOOGLE TANGO. 26

FIGURE 4-4 MICROSOFT HOLOLENS. 26

FIGURE 4-5 LEICA LASER SCAN. 26

FIGURE 4-6 3D RECONSTRUCTION FROM GoPRO IMAGES USING COLMAP (LEFT) AND LEICA SCANNER (RIGHT). 27

FIGURE 4-7 RECONSTRUCTION FROM DRONE IMAGES BY USING COLAMP (LEFT) AND AGISOFT PHOTOSCAN (RIGHT)..... 27

FIGURE 4-8 ALIGNED MULTI-ROOM USING GoPRO AND COLMAP. OUTSIDE (TOP) AND INSIDE (BOTTOM). 28

FIGURE 4-9 ALIGNED MULTI-ROOM USING GOOGLE TANGO. 28

FIGURE 4-10 ALIGNED MULTI-ROOM USING MICROSOFT HOLOLENS..... 29

FIGURE 4-11 GOOGLE TANGO DATA. *PROGRESS TRACKING* EXAMPLE, RED COLOUR INDICATES NEW CHANGES WITH RESPECT TO THE PREVIOUS CONSTRUCTION STAGE. MORE DETAILS IN DELIVERABLE D4.4..... 29

FIGURE 4-12 SCHEPENDOMLAAN DATASET. BIM FOR DIFFERENT WEEKS..... 30

FIGURE 4-13 3D RECONSTRUCTION OF WEEK 27 (LEFT) AND 30 (RIGHT) WITH COLMAP..... 30

FIGURE 4-14 BIM AND 3D RECONSTRUCTION. 31

FIGURE 4-15 BIM AND 3D RECONSTRUCTION. 32

FIGURE 4-16 ACQUA & SAPONE BUILDING. DURING CONSTRUCTION (ABOVE); FINISHED (BELOW)..... 33

FIGURE 4-17 BIM AND 3D RECONSTRUCTION FROM DRONE IMAGES ACQUISITION. 34

FIGURE 5-1 THE DATAFLOW OF THE VCMF AND BIM/BCF. 35

FIGURE 5-2 INTEGRATION WITH VCMF. THIS DIAGRAM INCLUDES THE WORK OF BOTH DELIVERABLES D4.3 AND D4.4. 37

FIGURE 5-3 EXAMPLE OF POINT CLOUD USE-CASE. 38

FIGURE 6-1 TAKE MEASUREMENTS IN MATTER OF SECONDS..... 40

FIGURE 6-2 INVERSE PROBLEM., FROM 3D POINT CLOUD TO BIM OR FLOOR PLAN. 41

FIGURE 6-3 SYSTEM OVERVIEW WITH THE MAJOR PROCESSING STEPS: 1) VOXEL GRID FILTERING REDUCES THE NUMBER OF 3D POINTS AND HOMOGENIZES THEIR DENSITY, 2) WALL DETECTION VIA PLANE CANDIDATE SELECTION 3) ROOM LAYOUT DETECTION VIA REASONING ABOUT DETECTED WALLS, AND 4) MODEL GENERATION..... 41

FIGURE 6-4 QUALITATIVE RESULTS OF OUR APPROACH ON FIVE INDOOR DATASETS. ALTHOUGH THE DATA IS MISSES SCANNED PARTS AND THE ROOMS CONTAINS A LOT OF FURNITURE, OUR METHOD RELIABLY COMPUTES THE ROOM LAYOUTS A VARIETY OF SCENES. 41

ABBREVIATIONS

B2S = Built to Specifications

DOA = Description of Action;

CS = Communication Strategy;

WP = Work Package.

SfM = Structure from Motion

SLAM = Simultaneous Location and Mapping

D4.3 = Deliverable 4.3

D4.4 = Deliverable 4.4

BIM = Building Information Model

IFC = Industry Foundation Classes

BFC = BIM Collaboration Format

VCMP = Virtual Construction Management Platform

1. INTRODUCTION

Capturing 3D models of buildings in any phase of their construction is an important cornerstone for subsequent storage, building information modelling (BIM), inspection, progress tracking as well as task management of 3D entities.

This document provides an overview of multiple 3D scanning devices and technologies that have been considered and developed for the Built2Spec project.

Outline: **Section 2** provides theoretical background information about different 3D presentations as well as some algorithms which are employed for the during the processing of 3D data. **Section 3** describes the different scanning technologies in more details and compares them with respect to their properties like: quality, accuracy, speed, and cost. **Section 0** provides examples of test datasets that have been acquired for the project. **Section 5** explains the data integration into the VCMP and BIMserver. **Section 6** describes, as additional work to D4.3, a method developed by ETH that automatically creates a BIM from previously scanned data. **Section 7** concludes the report.

Important note: Deliverable **D4.3** comprises the *data acquisition* with various technologies, while Deliverable **D4.4** (due at M36 – Dec. 2017) deals with *processing* and *analyzing* the 3D data acquired. Moreover, the content of this deliverable is also related to the Deliverables D2.2 (due at M29), D4.5 (due at M30), D7.2 (submitted at M28) and D7.3 (submitted at M28) to which we will refer for further details.

1.1 Overview

This section provides an outline of the processing pipeline for the 3D data acquisition which can be done with various sensors as well as the subsequent data analysis for progress tracking and accuracy benchmarking. The processing steps are listed in the following and are also depicted in Figure 1-1.

Processing steps:

1. Data acquisition (D4.3):

- By using **passive** sensors: i.e. images from human exploration or drones,
- By using **active** sensors: i.e. by using devices such as Google Tango, Microsoft HoloLens, or 3D Laser Scanners.

2. 3D Reconstruction pipeline (D4.4):

- Alignment and comparison against 4D BIM models,

3. Geometry Analysis (D4.4):

- Detect Discrepancies (positive or negative),
- Progress Tracking.

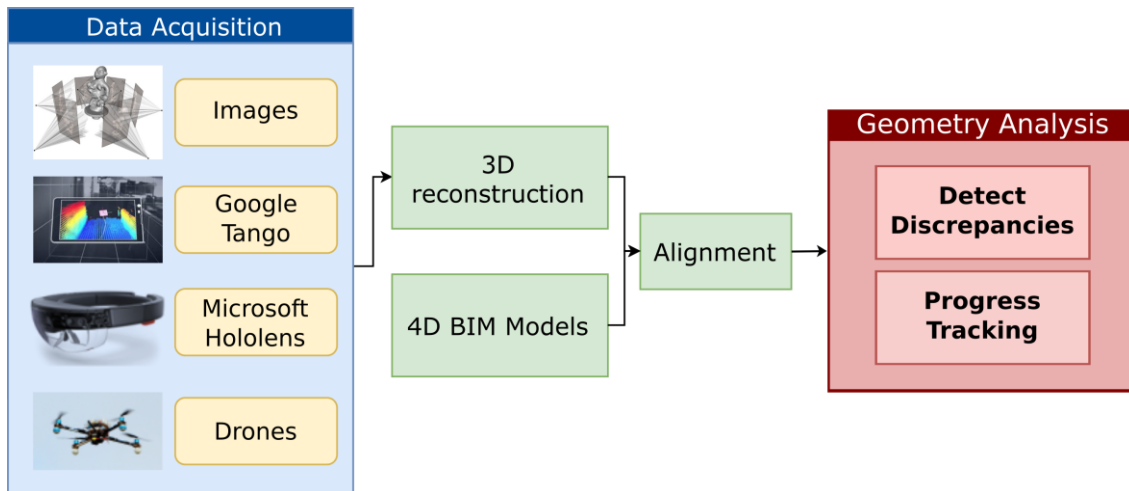


Figure 1-1 Overview of 3D data acquisition and data processing. Deliverable D4.3 comprises the data acquisition with various technologies. Deliverable D4.4 deals with processing and analysing the 3D data.

2. BACKGROUND

This section will introduce some basic concepts, which will help to understand the technologies behind the devices used to for the data acquisition described in this document. All the devices are mainly based on two techniques: **Structure from Motion (SfM)** and **Simultaneous Localization and Mapping (SLAM)**; described in more details in Section 2.2 and 0, respectively.

The goal of both SfM and SLAM is to estimate camera positions corresponding to a set of input images and simultaneously recovering a sparse 3D structure of the scene. Essentially, SfM and SLAM are solving a very similar problem, but while SfM is traditionally performed in an offline fashion, SLAM has been slowly moving towards the low-power / real-time / single RGB camera mode of operation. Many of the today’s top experts in Structure from Motion work for some of the world’s biggest tech companies, helping to make maps better. Successful mapping products like Google Maps could not have been built without intimate knowledge of multiple-view geometry, SfM, and SLAM. A typical SfM problem is the following: given a large collection of photos of a single outdoor structure (like the Coliseum, see Figure 2-1), construct a 3D model of the structure and determine the camera's poses.



Figure 2-1 Coliseum 3D reconstruction, from Flickr photo collection.

SfM is an essential component in this Deliverable, since it is used to obtain the Point Clouds (3D Reconstruction) from the images acquired with GoPro and Drones. In the other hand, SLAM is the method used in devices like Google Tango and Microsoft HoloLens. More details about these devices will be described later in Section 3.

2.1 Definitions

In order to describe the different 3D data acquisition technologies, we introduce some terminology in this section. To avoid repetition between deliverables, please refer to D2.2 for further literature review of techniques to perform camera calibration and automatic referencing.

2.1.1 Point cloud

A *point cloud* is a set of data points in a common coordinate system. Within the Built2Spec project, point clouds denote sets of 3D points that represent samples of captured surface of an object or a scene.

2.1.2 Mesh

A *mesh* is a collection of vertices, edges and faces in order to describe a surface in 3D space. The faces are described by polygons which connect neighboring vertices and are often restricted to triangles (*triangle mesh*) or quadrilaterals.

2.1.3 BIM

Building Information Modelling (BIM) is a process of creating and managing all the information during the project lifecycle, i.e. at planning, design, construction and operation phases of a physical infrastructure project. The output of this process is the Building Information Model, the digital description of every aspect of the built asset.

In the Figure 2-2 different shapes and 3D model representations are shown: point cloud (left), triangle mesh (middle), and BIM (right). The resolution of the mesh has been reduced to better visualize individual triangles.

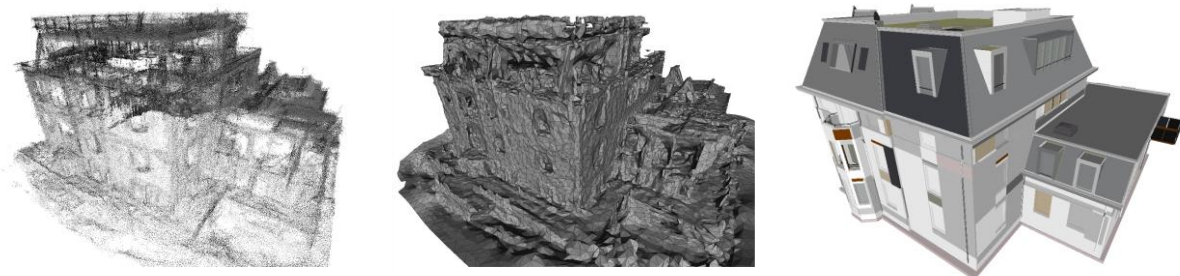


Figure 2-2 Model representations are shown: point cloud (left), triangle mesh (middle), and BIM (right).

2.2 Structure from Motion (SfM)

3D reconstruction from images traditionally first recovers a sparse representation of the scene and the camera poses of the input images using Structure from Motion (SfM). In our case, images

are obtained with GoPro cameras or using Drones taking pictures around an object, as it will explain later.

Summarized in Figure 2-3, Structure-from-Motion is the process of reconstructing 3D structure from its projections into a series of images. The input is a set of overlapping images of the same object, taken from different viewpoints (Figure 3-11). Ensure a proper level of visual overlap of the captured images is fundamental to achieve reasonable results when performing the 3D reconstruction.

The output is a 3D reconstruction of the object, and the reconstructed intrinsic and extrinsic camera parameters of all images.

Typically, Structure-from-Motion systems divide this process into the following stages:

1. Feature Extraction,
2. Feature Matching,
3. Geometric Verification,
4. Incremental Reconstruction.

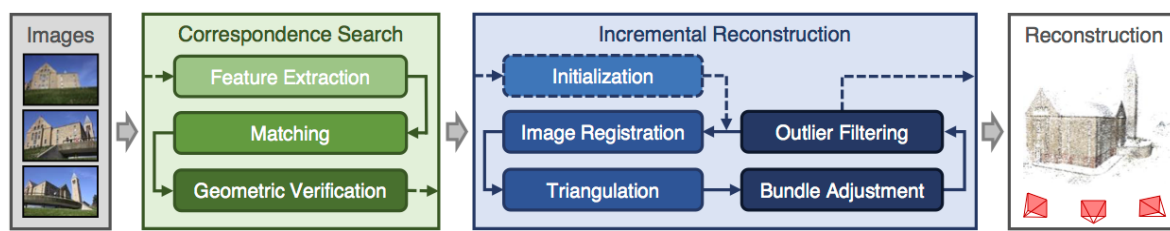


Figure 2-3 Overview of an incremental Structure-from-Motion pipeline.

To conclude this section, as general recommendations, to ensure optimal 3D reconstruction the following guidelines should be taken into account when capturing images:

- Choose a **good texture** avoiding completely texture-less images;
- Try to capture images at **similar illumination** conditions avoiding high dynamic range scenes and shiny surfaces.

Capture images with **high visual overlap** and from **different viewpoints**.

2.2.1 Feature Extraction

For each image, SfM pipeline detects a set of local features that are represented by an appearance descriptor. The features should ideally be invariant to radiometric and geometric changes, with the goal of being uniquely recognize in the matching stage in multiple images. For example, SIFT features [1] are highly distinctive, in the sense that a single feature can be correctly matched with high probability against a large database of features from many images.

2.2.2 Matching

Next, SfM discovers images that see the same scene part by leveraging the features extracted in the previous step. Since descriptors are high dimension vectors, a possible approach is to obtain the nearest neighbor of the scene from all model features in the descriptor space. A common method to this end is FLANN [2], which performs fast approximate nearest neighbor searches in high dimensional spaces.

2.2.3 Geometric Verification

The third stage verifies the potentially overlapping between image pairs. Since matching is based solely on appearance, it is not guaranteed that corresponding features actually map to the same scene point. Therefore, SfM verifies the matches by trying to estimate a transformation that maps feature points between images using projective geometry. Depending on the spatial configuration of an image pair, different mappings describe their geometric relation. A homography H describes the transformation of a purely rotating or a moving camera capturing a planar scene [3]. Epipolar geometry [3] describes the relation for a moving camera through the essential matrix E (calibrated) or the fundamental matrix F (uncalibrated), and can be extended to three views using the trifocal tensor [3]. If a valid transformation maps a sufficient number of features between the images, they are considered geometrically verified. Since the correspondences from matching are often outlier-contaminated, robust estimation techniques, such as RANSAC [4], are required. The output of this stage is a set of geometrically verified image pairs, their associated inlier correspondences, and optionally a description of their geometric relation.

2.2.4 Structure and motion reconstruction.

SfM initializes the model with a carefully selected two-view reconstruction. Choosing a suitable initial pair is critical, since the reconstruction may never recover from a bad initialization. Moreover, the robustness, accuracy, and performance of the reconstruction depends on the seed location of the incremental process. Initializing from a dense location in the image graph with many overlapping cameras typically results in a more robust and accurate reconstruction due to increased redundancy.

Image Registration. New images can be registered to the current model by solving the Perspective-n-Point (PnP) problem [5] using feature correspondences to triangulated points in already registered images (2D-3D correspondences).

Triangulation. A newly registered image must observe existing scene points. In addition, it may also increase scene coverage by extending the set of points through triangulation [6]. This is a crucial step in SfM, as it increases the stability of the existing model through redundancy [7] and it enables registration of new images by providing additional 2D-3D correspondences.

Bundle Adjustment (BA). Image registration and triangulation are separate procedures, even though their products are highly correlated - uncertainties in the camera pose propagate to triangulated points and vice versa, and additional triangulations may improve the initial camera pose through increased redundancy. Without further refinement, SfM usually drifts quickly to a non-recoverable state. BA is the joint non-linear refinement of camera parameters and point parameters that minimizes the re-projection error, in order to mitigate errors. See more information in [7] and Section 3.2.

2.3 Simultaneous Localization and Mapping (SLAM)

SLAM is a real-time version of Structure from Motion. Visual SLAM or vision-based SLAM is a camera-only variant of SLAM which forgoes expensive laser sensors and inertial measurement units (IMUs). Monocular SLAM uses a single camera while non-monocular SLAM typically uses a pre-calibrated fixed-baseline stereo camera rig.

A multitude of combinations of sensing modalities have been used for SLAM, including LIDAR, RGB-D cameras, inertial sensors, GPS, and vision. LIDAR is a form of active sensing, which consists of measuring laser time of flight to infer depth in real time. RGB-D cameras (structure light) projects a narrow band of light onto a three-dimensionally shaped surface produces a line of illumination that appears distorted from other perspectives than that of the projector, and can be used for an exact geometric reconstruction of the surface shape.

SLAM is effectively a special case of the SfM problem, in that it is specific to robotics and augmented/virtual reality applications, yet retains many of the same difficulties plus the additional constraint that it should be in real-time; making it an extremely challenging problem. Google Tango and Microsoft HoloLens have their own implementation of a real-time localization systems, a variant of Visual-Inertial SLAM using RGB-D cameras (see [8] for more information).

3. TECHNOLOGY DESCRIPTION

3.1 Devices Description

In the following sections, an overview of the technologies used during the B2S project is presented. We briefly describe the technologies and capturing devices which have been used for obtaining 3D reconstructions. In particular, these are the COLMAP's Structure-from-Motion pipeline applied to videos captured with a GoPro camera, Drones, Google Tango Tablet, Microsoft HoloLens, and 3D Laser Scan.

The attention in this section is posed on *“how each technology works”* and *“how is it possible to obtain the 3D Point Cloud”* starting from survey. Further details about *“how to do”* can be found in **D7.3 - Long Term Testing Teams Workshop materials** (submitted at M28).

3.1.1 GoPro

GoPro cameras has been used, see Figure 3-1, since it has a fish-eye lens which can capture a wide field of view and, in record mode, it has a high frequency rate (above 60Hz), allowing to record videos without blur. Therefore, some frames can be extracted as post-processing from these videos instead of continuously take pictures that can be time consuming.



Figure 3-1 GoPro Hero.

As general recommendations, if the picture capture process is controlled over, please follow these guidelines for optimal reconstruction results:

- Capture images with **good texture**. Avoid completely texture-less images (e.g., a white wall or empty desk). If the scene does not contain enough texture itself, you could place additional background objects, such as posters, etc. Good texture provides unique features on the object surface which is essential for matching them across different images and estimating their 3D position.
- Capture images at **similar illumination** conditions. Avoid high dynamic range scenes (e.g., pictures against the sun with shadows or pictures through doors/windows). Avoid specularities on shiny surfaces. The feature matching across images becomes significantly harder with strongly varying illumination and correspondences might not be found.
- Capture images with **high visual overlap**. Make sure that each object is seen in at least 3 images – the more images the better. Overlap ensures that many visual features are available for matching features and subsequently estimating the relative camera motion.
- Capture images from **different viewpoints**. Do not take images from the same location by only rotating the camera, e.g., make a few steps after each shot. At the same time, try to have enough images from a relatively similar viewpoint. Different viewpoints are important to see the same surface several times to estimate feature locations in 3D and to reduce the amount of unobserved areas due to occlusions. At least 3 to 5 images depicting the surface features should be taken for reasonable results.

3.1.2 Drone (EURECAT & R2M)

Drone survey allows a rapid 3D-reconstruction of the external geometry of the building simply starting from the post-processing of the acquired pictures.

The main advantages of this technology are here below listed:

- high-quality and low-cost data: one can get an area survey just in few minutes of fly allowing the acquisition of high value information and high accuracy pictures (about 2 cm precision at ground);

improved safety: no need to deploy personnel in hazardous areas without any health risks for the surveyors working on site. Drone survey and related imageries data post-process technology brought to the market a very new powerful “tool” to survey your construction site reducing the time spent collecting accurate data - by acquiring imageries from the flight you can gather millions of high precision data points per flight improving also the safety since no personnel is employed directly in hazardous areas. In other words, drone makes the data collection simpler so you can focus your energy on using and analyzing the acquired data, rather than working out how to gather it.

Anyways the dataset, i.e. the images to be captured by drone on site, is the key aspect in order to generate a good 3D reconstruction from images with high quality and accuracy. This dataset acquisition depends on the following key points:

- High overlap between images;
- Image acquisition plan, which depends on the type of terrain/object to be reconstructed, which in turn depends on:
 - Path planning,
 - Flight height,
 - Camera angle(s),

- Number of flights;
- Ground Control Points (GCPs) to improve and/or validate the accuracy of the geo-referencing.

However, the sensor selection must be done before the definition of the image acquisition plan. In this sense, there is a trade-off between the shutter speed, the aperture and the ISO sensitivity of the camera. For processing, the images should be sharp and have the least amount of noise. As a rule of thumb:

- **Shutter speed** should be fixed and set to medium speed (between 1/300 second and 1/800 second), but fast enough to not produce blurry images.
- **ISO** should be set as low as possible (minimum 100). High ISO settings introduce noise and reduce the quality of the results.
- **Aperture** depends on the lens and it is better to leave it on automatic.

Once the camera has been selected, the design of the image acquisition plan can be started. The first step is to decide the Ground Sampling Distance (GSD), which is the distance between the centers of two consecutive pixels on the ground; therefore, it is directly related to the resolution of the 3D reconstruction. By minimizing the GSD, the accuracy and the quality of the final results will be better as well as the number of details that are visible in the final 3D reconstruction will be higher. Thence, minimize the GSD must be a priority for Built2Spec purposes.

The flight height that is needed to obtain a given GSD can be computed and depends on some parameters of the camera. The following equation shows the relation between the flight height, the GSD and the camera parameters:

$$H[m] = \frac{imW * GSD * F_R}{S_W * 100}$$

where:

- $H[m]$ is the flight height;
- imW is the image width in pixels;
- GSD is the ground sampling distance;
- F_R is the real focal length;
- S_W is the real sensor width.

The final step is the definition of the path planning, which depends on the terrain/object to be reconstructed. The following pictures show the common path plans for different cases.

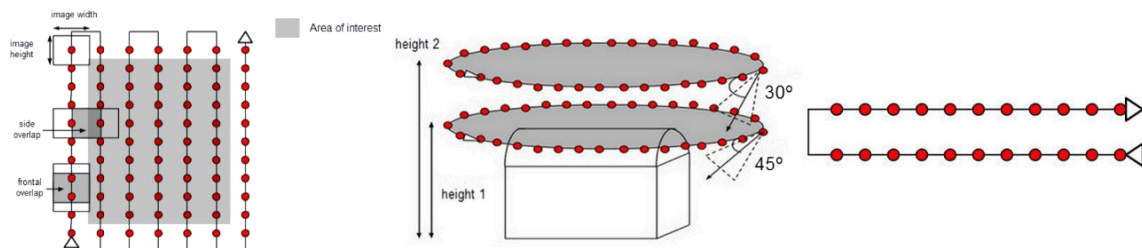


Figure 3-2 Path plans for different cases. From left to right. Building reconstruction and corridor mapping.

In a general case, it is recommended to take the images with a regular grid pattern and use an overlap of 75% with respect to the flight direction and at least 60% between flying tracks (Figure 3-3).

For a building reconstruction, it is recommended to do multiple flights around it at different altitudes and camera angle (as it is shown in Figure 3-2). It is recommended to take one image every 5-10 degrees to ensure enough overlap.

Finally, a structure similar to a corridor (road, bridge, viaduct, etc.) requires at least two flight lines with more frontal overlap (at least 85% - see Figure 3-2). However, most of the structures do not fit completely in one of the presented cases. Then, multiple flights must be done to ensure the completeness of the reconstruction.

The following points must be taken in to account to do multiple flights:

- The different plans must be taken as much as possible under the same conditions (sun direction, weather conditions, etc.).
- There must be enough overlap between the different image acquisition plans (Figure 3-3).
- The flight height should not be too different between the flights.

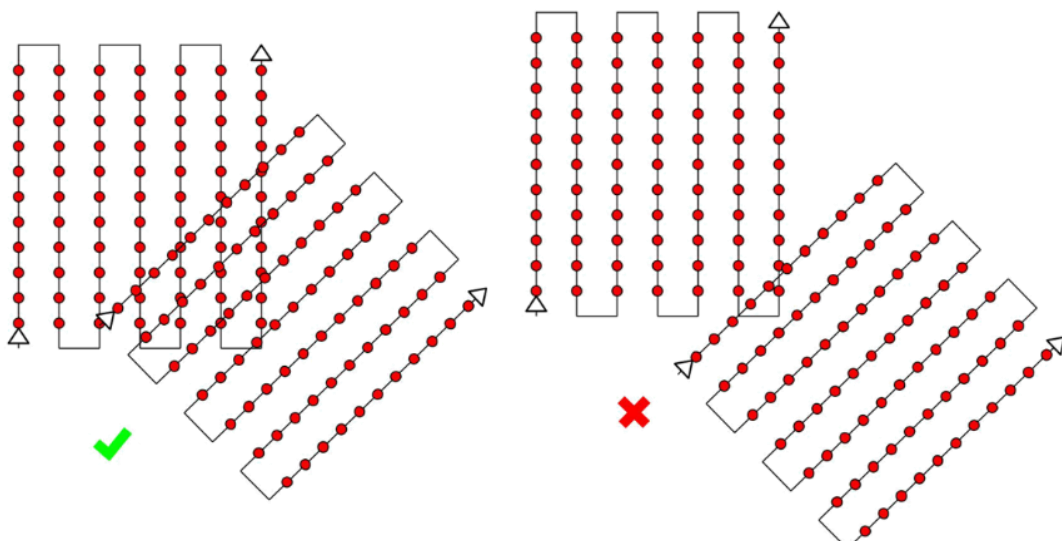


Figure 3-3 Overlap between different image acquisition plans, with a right configuration on the left and an incorrect one on the right.

The last point before flying is the distribution of the **Ground Control Point (GCP)**. A GCP is a landmark clearly visible in the aerial images, which has been precisely georeferenced. They must be placed homogeneously in the area of interest. It is also recommended to place one GCP in the center of the area in order to further increase the quality of the reconstruction. A minimum number of three GCPs is required to be useful for reconstruction.

As an example, Eurecat defined an image acquisition plan over a viaduct, which ensured the completeness of the reconstruction. The image below represents the different flights and the location of the GCPs.

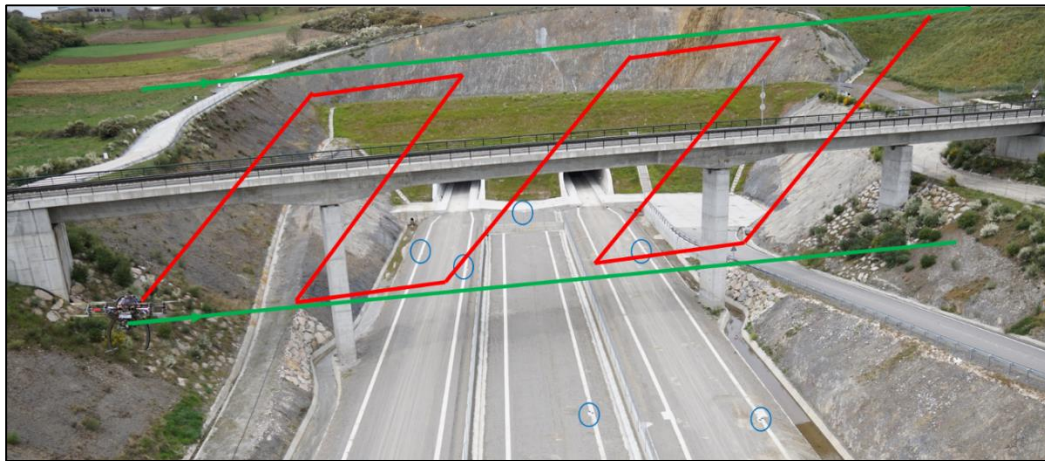


Figure 3-4 Image acquisition plan designed by Eurecat for the reconstruction of the viaduct. A grid pattern was used for the first flight (red line), and two flights were done at the frontal and rear side (green lines).

In this scenario, the Eurecat’s drone did three flights. A grid pattern at 40m (with a GSD of 0.5cm) was used for the first flight (red line in the image above), and two flights were done at the frontal and rear side of the viaduct at 30m (green lines in the image above). Moreover, seven GCP were distributed homogeneously in the area (six of them are shown with blue circles in the image above).

The results of the image acquisition plan were also improved by using a Real Time Kinematic satellite navigation technique (RTK) which enhance the precision of the position data by using a reference station, which provides real-time corrections providing up to centimeter-level accuracy. RTK was used for the navigation, on the other hand a Post Processed Kinematic method (PPK) was used to geo-reference the images. PPK is a method, which processes the positioning information after the flight. Data is logged in the drone and combined with the data from the base station when the flight is completed. As a result, there is no risk of data loss due to radio links outages. Therefore, PPK is a little more thorough than RTK.

The following image shows the result of the 3D reconstruction:



Figure 3-5 3D reconstruction of the viaduct done by Eurecat (Pix4D, left) and by ETH (COLMAP, right).

3.1.3 Google Tango

Google has invested in the development of a tablet devices (Figure 3-6) that can scan and build 3D models of their environments. ETH is directly involved in the called Project Tango and the application to capture 3D is described in the Figure 3-7.



Figure 3-6 Google Tango Tablet.

The Google Tango device is constantly tracking its location and orientation by evaluating measurement from the Inertial Measurement Unit (IMU) and by matching of feature points found in the input images. With this tracking the depth maps from the internal depth sensor can be registered into a global coordinate system and a dense map of the environment is created and grown over time. The dense 3D map computation of Project Tango [8], [9] is based on Truncated Signed Distance Function (TSDF) fusion [10] of depth maps into a voxel volume via voxel block hashing [11]. Then a surface triangle mesh is extracted as an iso-surface of the TSDF corresponding using the Marching Cubes algorithm [12]. It gives a reconstructed mesh and the camera pose. Lastly, a point cloud can be obtained by extracting the mesh vertices or by point sampling on the mesh surface.

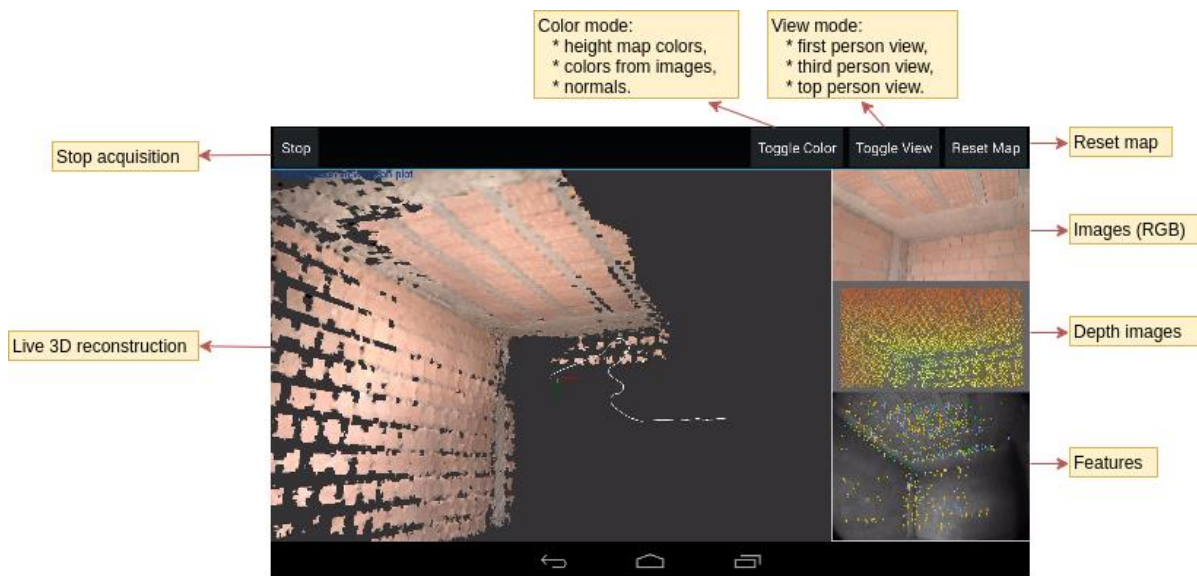


Figure 3-7 Google Tango application: Tango Mapper.

3.1.4 Microsoft HoloLens

The HoloLens is a head-mounted display unit connected to an adjustable, cushioned inner headband, which can tilt HoloLens up and down, as well as forward and backward. To wear the unit, the user fits the HoloLens on their head, using an adjustment wheel at the back of the headband to secure it around the crown, supporting and distributing the weight of the unit equally for comfort, before tilting the visor towards the front of the eyes. The tracking, and mapping technology of Microsoft HoloLens is proprietary and has not been disclosed to the public. However, the input sensors as well as the internal processing pipeline for camera tracking and dense mapping is very similar to the one from Google Tango (refer to section 3.1.3 for dense mapping).



Figure 3-8 Microsoft HoloLens.

3.1.5 Leica Laser Scanner

A P40 Leica 3D laser scanner is a device which scans a target area to extract a point cloud (a cloud of points with known coordinates). It is a technology well known in the market and the main advantages are its relatively portable nature and its resolution/time ratio.

P40 Leica 3D laser scanner can perform in temperatures ranging from -20°C to $+50^{\circ}\text{C}$ and comply with the IP54 rating for dust and water resistance.

P40 Leica 3D laser scanner can deliver:

- High quality 3D data and HDR imaging
- Extremely high speed scan rate of 1 mio points per second
- Ranges of up to 270 m
- Low range noise
- Survey-grade dual-axis compensation
- Highly detailed 3D colour point clouds mapped in realistic clarity

Technical information about the scanner can be found at:

http://hds.leica-geosystems.com/downloads123/hds/hds/general/brochures-datasheet/Leica_ScanStation_P30-P40_BIM_DS_en.pdf



Figure 3-9 Leica Laser Scanner.

3.1.6 Comparison of Devices

Capturing images with a consumer camera like a GoPro is simple, intuitive and fast both when capturing indoor and outdoor environments. However, for subsequent image-based 3D reconstruction, some experience is required to aid the reconstruction algorithm by considering the

right distances and angles to object surfaces, light conditions and explore settings as well as sufficient image overlap between images.

With the help of a drone, outdoor building scans can be captured in little time and it is hence the best device for capturing outdoor properties from all possible viewpoints with high coverage. In good conditions with respect to light and scene surface properties, as well as image quality and sufficient image coverage high quality, 3D models can be computed that can compete with accuracy of a laser scanner, but at a fraction of the costs for both hardware expenses and acquisition time. Anyway, some experience is needed to deal with software/algorithms 3D reconstruction as per GoPro.

Laser scanning is the expensive gold standard and remains necessary in textureless scenes or for benchmarking or completing other scanning modalities.

Microsoft HoloLens and Google Tango are useful technologies for indoor real-time building reconstruction which provide immediate user feedback with respect to capturing quality and scene coverage also allowing for instant data usage after the acquisition. In comparison, for the pure scanning task, the Tango tablet is more comfortable to use when a dense 3D model is required, because the depth camera needs to be moved around to various viewpoints to ensure full coverage without occlusion holes. In contrast, HoloLens provides a more immersive experience for augmented reality applications like life onsite inspections for which only the current view point is important, but a highly detailed and complete 3D scan of the scene is not necessary. In contrast to the Tango processing pipeline, HoloLens does currently not allow to save all input data for later offline post-processing to obtain higher quality models.

The following table provides an overview of scanning devices with respect to purchase costs, weight, required training time, suitability for indoor and outdoor scanning as well as usage restrictions.

	Device cost	Weight/portability HIGH-MEDIUM-LOW	Training time to learn	Indoor / Outdoor	Particular restrictions
GoPro	\$ 500	HIGH (88 g)	1-2 days	Both	--
Drone	The cost range from 3K€ up to 15k€ or more for a professional drone. Along with the device cost one has to take into account the cost of the licence (about 2000€) and the cost of the insurance (about 400€/year)	MEDIUM/LOW The drone is transported in a rigid box (50x50 cm h=50 cm). Its weight is about 12 Kg	Pilot must be licensed and acquire good practice.	Outdoor	License is required. There are particular "no flight zones" according to National Flight Regulation
Google Tango	\$ 500	HIGH (370 g)	1 day	Indoor	--
Microsoft HoloLens	\$ 3000	HIGH (579 g)	1 day	Indoor	--
3D Leica scanner	€35,000 (exc. VAT)	LOW portability ~23 kg without a tripod (scanner + batteries = 15 kg). Case Dimensions: 500 mm x 640 mm x 370 mm	1 day	Both	With outdoor scans: rain / strong winds

3.2 3D Reconstruction Software Description

An overview of the software employed to reconstruct a 3D Point Cloud from a set of input photographs is presented in the following.

3.2.1 COLMAP (ETH)

COLMAP [13], [14] is a general-purpose Structure-from-Motion (SfM) and Multi-View Stereo (MVS) pipeline with a graphical and command-line interface. It offers a wide range of features for reconstruction of ordered and unordered image collections. The software is licensed under the GNU General Public License (open-source), and is maintained by a member of ETH.

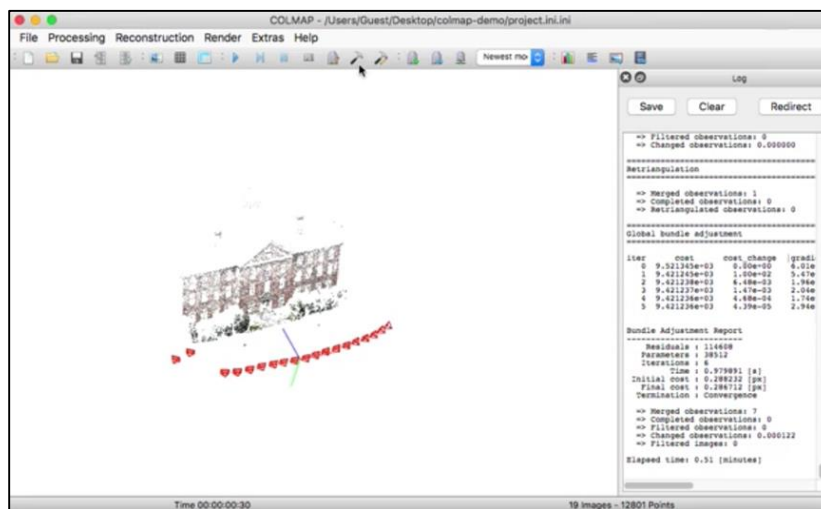


Figure 3-10 COLMAP's interface.

3.2.2 Agisoft Photoscan

Agisoft PhotoScan is a professional stand-alone software developed by *Agisoft LLC* located in St. Petersburg in Russia for a photogrammetry pipeline (see Figure 3-11). The software is available in:

- **Standard licence:** is a basic program realization that, however, could be used to solve various tasks: reconstruction, modelling, digitalization of objects and scenes as well as in video games creation, etc. Main Features:
 - Automatic camera calibration;
 - Automatic tie-point search;
 - Aerial triangulation and block adjustment;
 - 3D model generation and export in TIN formats.
- **Pro licence:** includes all Standard edition functionality, that is enhanced with the following features:
 - Georeferencing;
 - DEM export in GeoTIFF elevation data, Arc/Inpho ASCII grid, Band interleaved file format, XYZ file formats;
 - Orthophoto generation;

- Carrying out area and volume measurements;
- Python scripting.

This software performs photogrammetric processing digital images and generates 3D Point Cloud, visual effects production as well as for indirect measurements of objects of various scales., the standard version is sufficient for interactive media tasks, while the Pro version is designed for authoring GIS content. The software is widely used by many UAV (drone) companies.

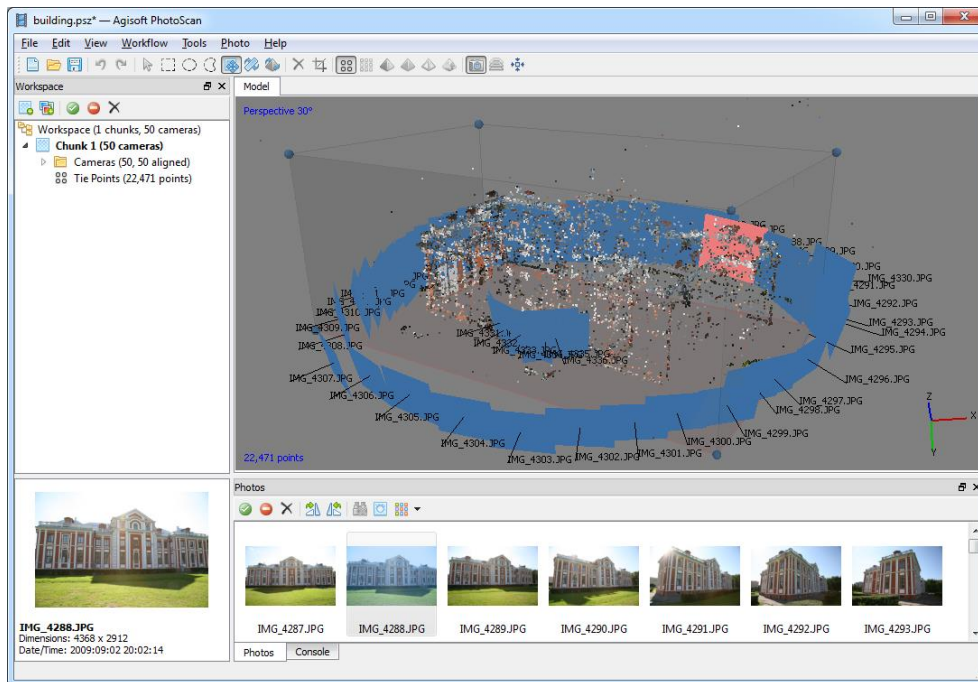


Figure 3-11 Overlapping and viewpoints (from Agisoft Photoscan guideline).

The general workflow to reconstruct a 3D Point Cloud can so summarized:

- 1) **Source Data: Drone Survey images;**
- 2) **Step 1: Image Management:** this step is aimed to input the data into the software and to determine the structure of the project, thus having everything prepared for the processing procedure to start. The raw data can be added to PhotoScan in the form of digital imagery in one of the following formats: TIFF, DNG, JPEG, BMP, PNG, PPM and JPEG MPO. To cope with huge data sets (>10,000 photos) PhotoScan suggests "chunk" conception, which enables to process the data in parts merging the results later. At this stage of the workflow, one can assign a camera model to each image (i.e. focal length in x and y dimensions measured in pixels, principal point coordinates, radial and tangential distortion of the lens and skew transformation coefficients)
- 3) **Step 2: Ground Control. Aerial Triangulation:** to strengthen overall geometric robustness of the image block and to allow for more precise estimation of the exterior orientation parameters, which improves georeferencing accuracy of the final results, it is useful to add Ground Control Points (GCPs) into the project – GCPs are simply targets with known/measured positions. PhotoScan can process GCPs coordinates recorded in various coordinates systems. Once GCPs have been added to the project, it is the time to run aerial

triangulation procedure. The resulting output of this step, is the numerical estimate of how well all the data fits each other – i.e. spatial and georeferencing accuracy.

- 4) **Step 3: Dense Surface Reconstruction:** this step deals with the Digital Elevation Model (DEM) generation setting effective output resolution cropping invalid regions as well as defining the region of the reconstructed area to be included in the resulting DEM. PhotoScan supports DEM export as GeoTIFF elevation data, Arc/Inpho ASCII grid, XYZ and Band interleaved file formats. Furthermore, 3D model formats like OBJ, 3DS file format, VRML, COLLADA, Stanford PLY, Autodesk DXF, U3D, Adobe PDF can be exported.
- 5) **Step 4: Orthophoto Generation – *additional*:** once created the 3D model of the surveyed PhotoScan allows the generation of a very accurate orthophoto. The orthophoto is reliable and can be either used as a final-end product or as a basis for GIS data for specific purposes creation in external tools.

3.2.3 Pix4D

Pix4Dmapper is a commercial photogrammetry software developed by Pix4D. It uses images to generate point clouds, digital surface and terrain models, orthomosaics, textured models and more (see **Figure 3-12**). The software is available in:

- **Yearly/monthly licence.** Main features:
 - Rent for 1 year/30 days;
 - Unlimited processing;
 - Desktop + Cloud;
 - Support & upgrade included.
- **Perpetual licence.** Main features:
 - Perpetual desktop software;
 - Unlimited processing;
 - Desktop + Cloud;
 - Support & upgrade included for 1 year.

The workflow, very similar to the Agisoft Photoscan one (see section 3.2.2), to reconstruct a 3D Point Cloud can so summarized:

- 1) **Step 1: Image Management:** add photos and select the pictures you want to post-process;
- 2) **Step 2: Calibration:** overlapping pictures;
- 3) **Step 3: Dense Surface Reconstruction:** point cloud – orthomosaics model.

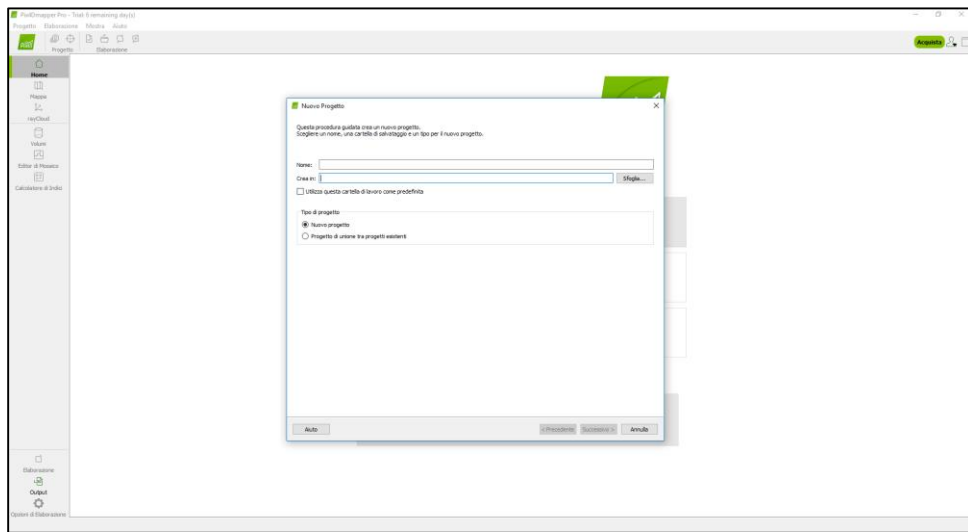


Figure 3-12 Pix4D Mapper interface

3.2.4 Comparison of 3D Reconstruction Software

The following table provides a brief overview of the 3D reconstruction software packages considered in the project and lists their most important properties.

	License cost	Inputs	Supported format output	Reference
COLMAP	GNU GPL (Opensource)	Images (JPEG)	Point Cloud and Mesh (PLY)	https://colmap.github.io/
Pix4D	Subscription: <ul style="list-style-type: none"> Yearly = 2.6 k€ Monthly = 260 € Perpetual: 6.5 k€	*.JPEG; VIDEOS	Main output: *.PLY, *.LAS, *.LAZ, *.XYZ Others: Flythrough video; Orthomosaic; DSM, DTM, 3D textured model, 2D measurement, 2D vector output.	https://cloud.pix4d.com
Agisoft Photoscan	Perpetual: <ul style="list-style-type: none"> Agisoft PhotoScan PRO - stand-alone license = 3.5 k\$ USD Agisoft PhotoScan Standard Edition - stand-alone license = 179 \$ USD 	*.JPEG	Main output: *.PLY, *.LAS, *.OBJ, *.XYZ Others: Orthomosaic; DEM; 3D textured model.	http://www.agisoft.com

4. REAL DATA ACQUISITION

In this section different case studies are presented in order to show all the solution tested based on the development in T4.3 and T4.4 activities. More details about the result will be presented in D4.4.

4.1 Italian demo site “San Salvo” Dataset (R2M & DE5)

The dataset was acquired within two days. The demo site, placed in San Salvo, Italy (Abruzzo region), is under construction. It is a residential building with 2 underground floors and 4 above ground floors. The house contains 11 apartments and 20 garages. Further details are given in deliverable D7.2. The following subsections show images of scan results with a variety of devices for both indoor and outdoor scans.



Figure 4-1 San Salvo - Italian demo site. BIM (left) and north-east view (right).

4.1.1 Indoors

The following images show single room indoor views in the San Salvo construction site, acquired with different device modalities, in order to establish a qualitative comparison between them. See Figure 4-2 (GoPro + COLMAP), Figure 4-3 (Google Tango), Figure 4-4 (Microsoft HoloLens), and Figure 4-5 (Leica Laser Scan).



Figure 4-2 GoPro + COLMAP: 3D Reconstruction from GoPro images after processing with COLMAP.

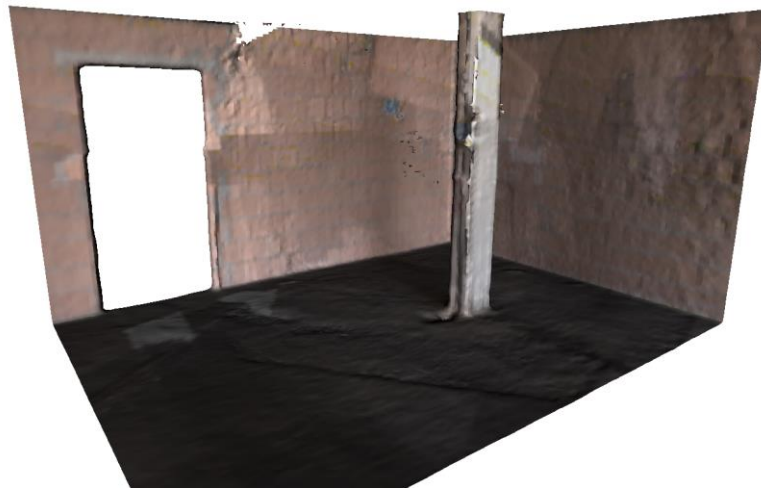


Figure 4-3 Google Tango.

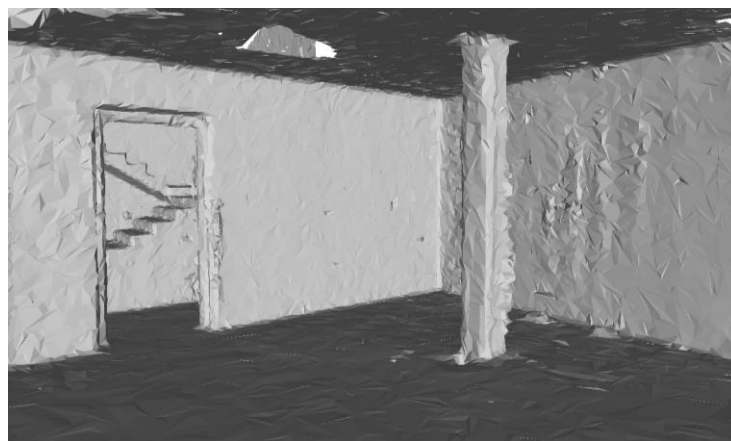


Figure 4-4 Microsoft HoloLens.

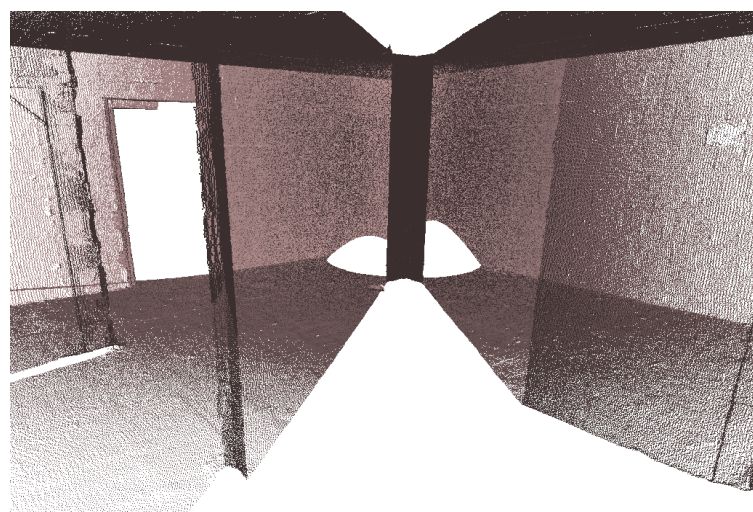


Figure 4-5 Leica Laser Scan.

4.1.2 Outdoors

Due to their small depth range and sensitivity to direct sunlight the active depth sensors used in the Google Tango and the HoloLens devices are not suited for outdoor data acquisition. In this application Drone, Leica and GoPro devices have been used to acquire data.

The following figures give an overview on the level of detail that can be achieved outdoors with the other modalities considered here. Figure 4-6 shows a reconstruction with COLMAP from GoPro images, next to the result with the Leica laser scanner. Figure 4-7 compares the results of 3D reconstructions with COLMAP and Agisoft Photoscan software, both run with the same image set that has been captured with a drone.



Figure 4-6 3D Reconstruction from GoPro images using COLMAP (left) and Leica Scanner (right).

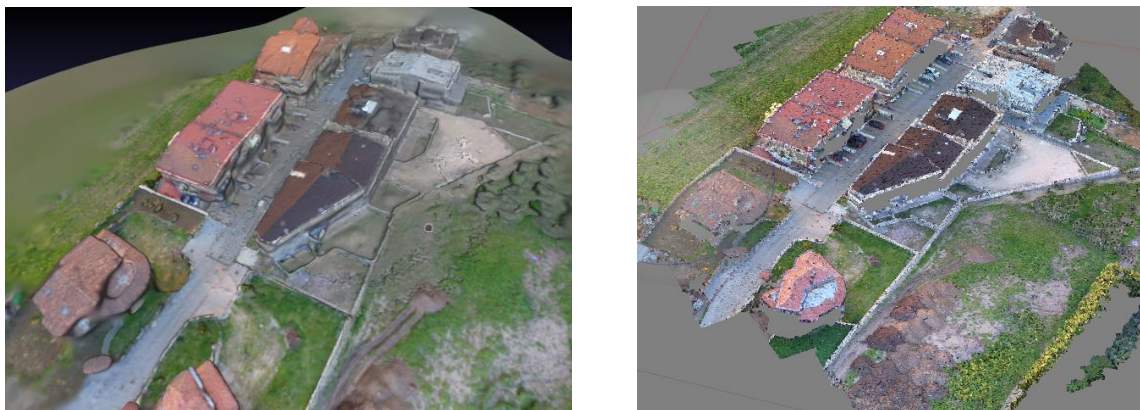


Figure 4-7 Reconstruction from Drone images by using COLAMP (left) and Agisoft Photoscan (right).

4.1.3 The results achieved from the different devices employed are in line with the expected results. Multiple rooms

Multiple rooms can be acquired individually and subsequently merged later in an alignment process. A merged model from different rooms using GoPro can be seen in Figure 4-8 (including inside and outside information). **Note:** the alignment problem is a research topic studied in the Deliverable D4.4.

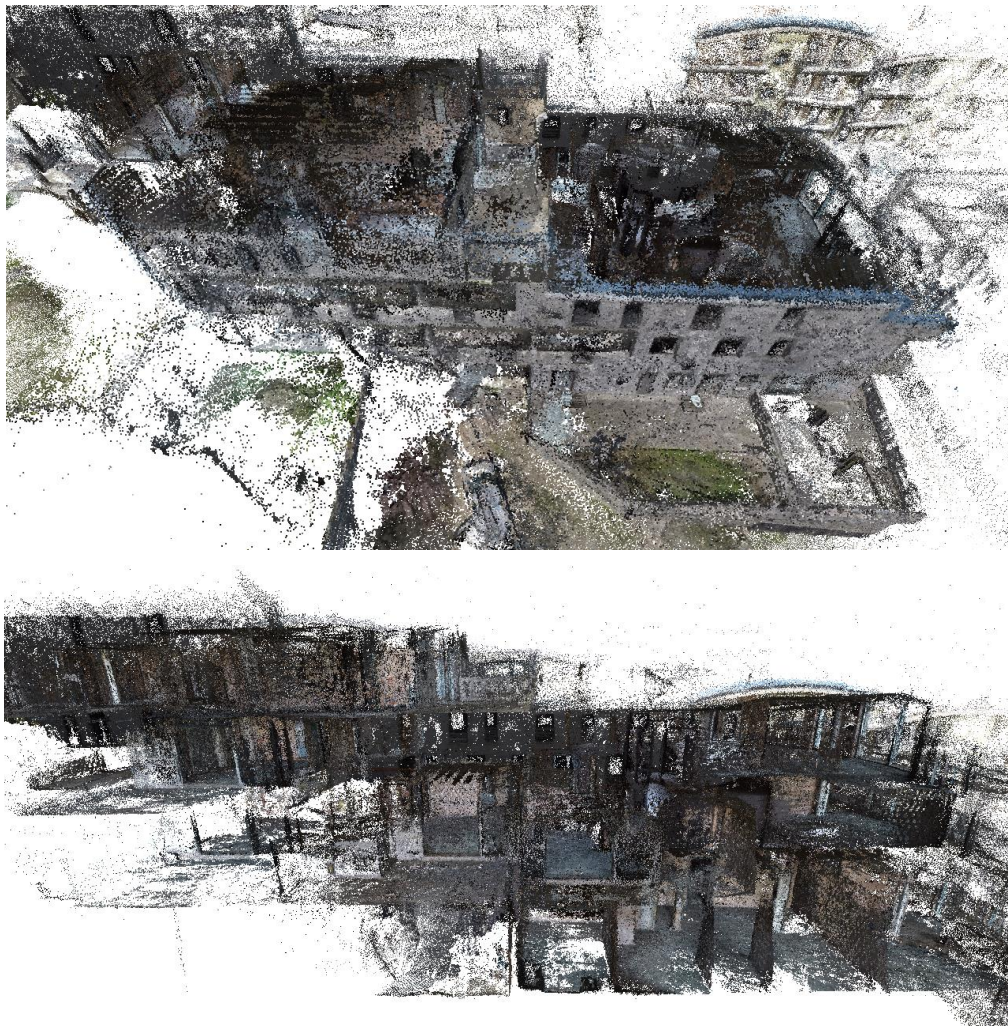


Figure 4-8 Aligned multi-room using GoPro and COLMAP. Outside (top) and Inside (bottom).

Similar results can be obtained with Google Tango (Figure 4-9) and Microsoft HoloLens (Figure 4-10), but in these cases only show inside information since their active sensors (based on laser projection patterns) do not perform outdoors.

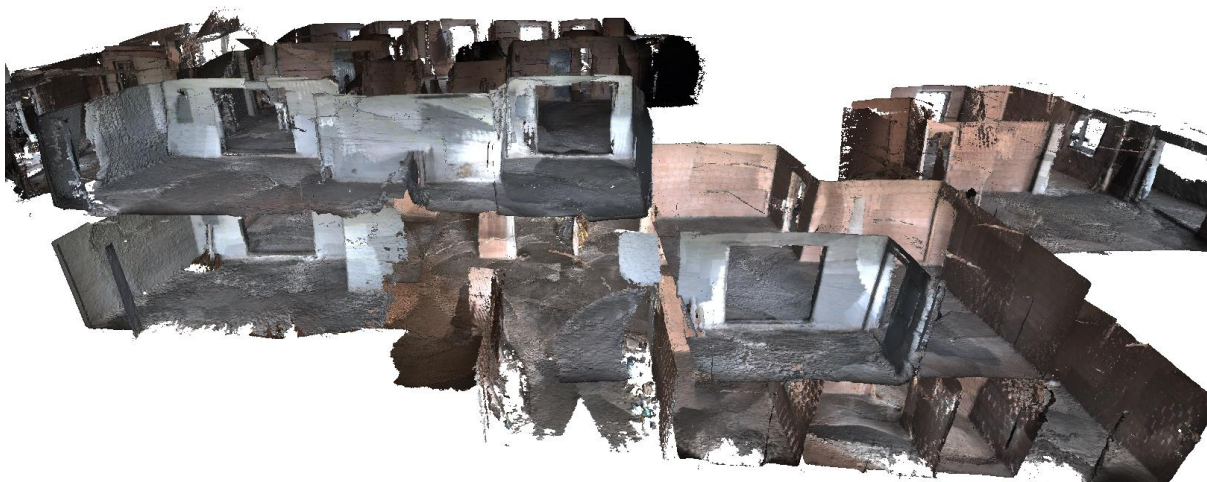


Figure 4-9 Aligned multi-room using Google Tango.

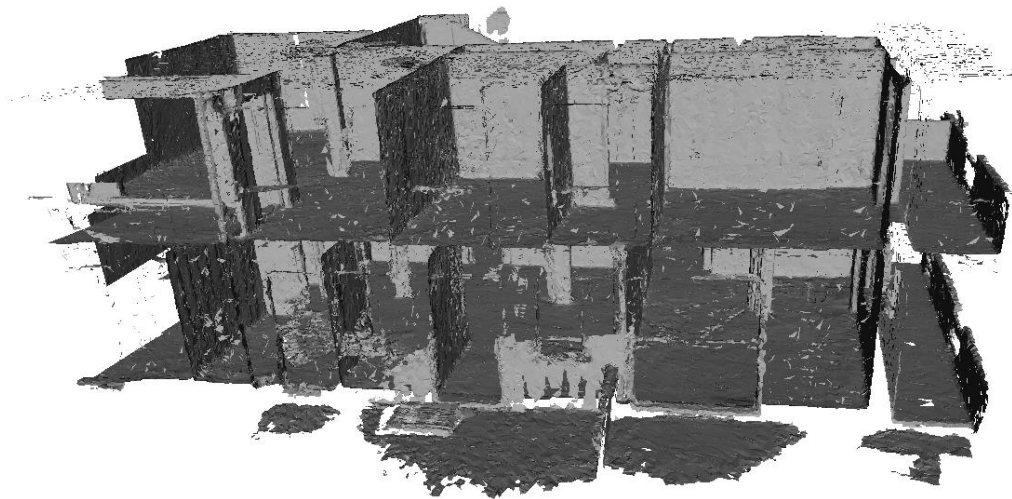


Figure 4-10 Aligned multi-room using Microsoft HoloLens.

4.1.4 Progress Tracking

Data was acquired during three construction stages, shown in Figure 4-11 as **Level 0** (initial stage), **Level 1** (intermediate stage), and **Level 2** (final stage). Figure 4-11 also shows the difference between stages using a colour map, indicating which area has built changed (red colour) with respect to previous stage. This **Progress Tracking** technique can be useful to identify if the construction is being built up to the specifications (as the name of the project indicates: Built2Spec). Further information will be presented in Deliverable D4.4.

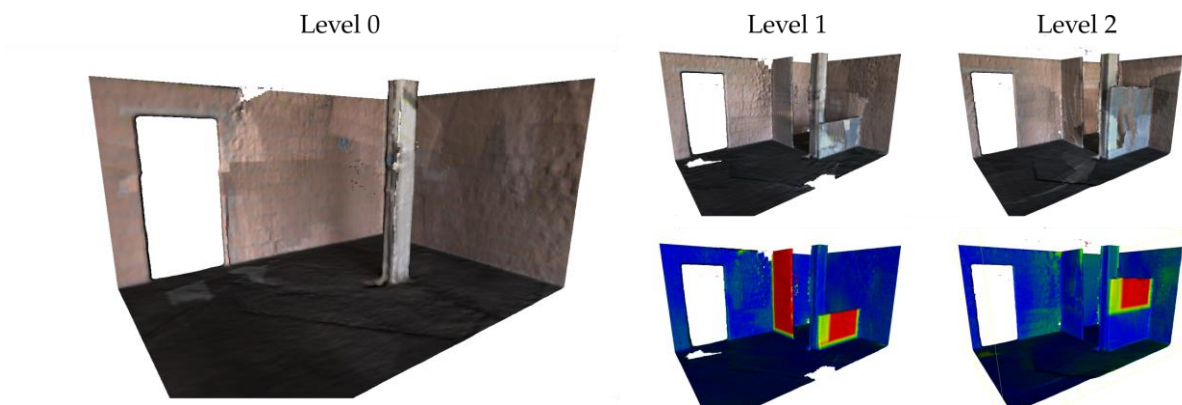


Figure 4-11 Google Tango data. *Progress tracking* example, red colour indicates new changes with respect to the previous construction stage. More details in Deliverable D4.4.

4.2 Dataset Schependomlaan (TNO)

The dataset contains the following elements:

- Design model in .IFC and .PLA (Archicad)
- Issues (collision / clash detection) in BCF (.bzfzip) and in Tekla BIMsight Package.
- Subcontractor models in .IFC and .DWG - Flooring - Walls - Stairs - Fencing - Steel - Roofs - Prefab
- Coordination models in .TBP (Tekla BIMsight Package)
- Schedule/Planning in .pdf and .xml

- As-planned models in .IFC and Synchro file format.
- As-built models in point cloud formats .ASCII and .PLY
- Results comparison as-planned and as-built models in .xls
- As-planned Event log in .xlsx and .csv
- As-built Event log in .xlsx and .csv
- Event log with actors in .xlsx and .csv
- Download link to drone images and videos

Detailed information about this dataset can be found in the following link:

<https://github.com/openBIMstandards/DataSetSchependomlaan>

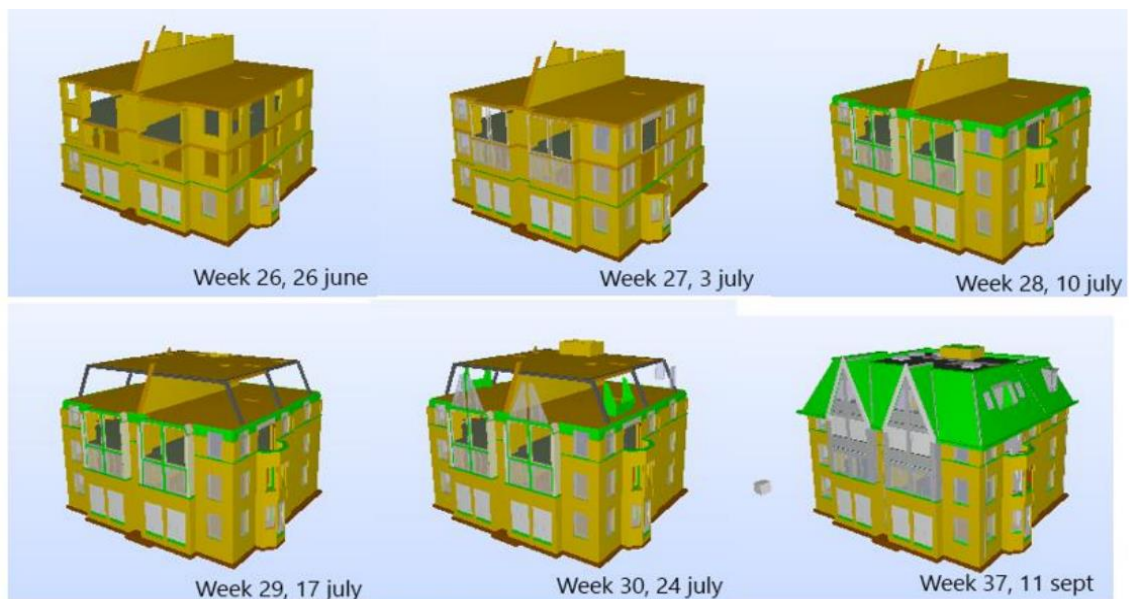


Figure 4-12 Schependomlaan dataset. BIM for different weeks.

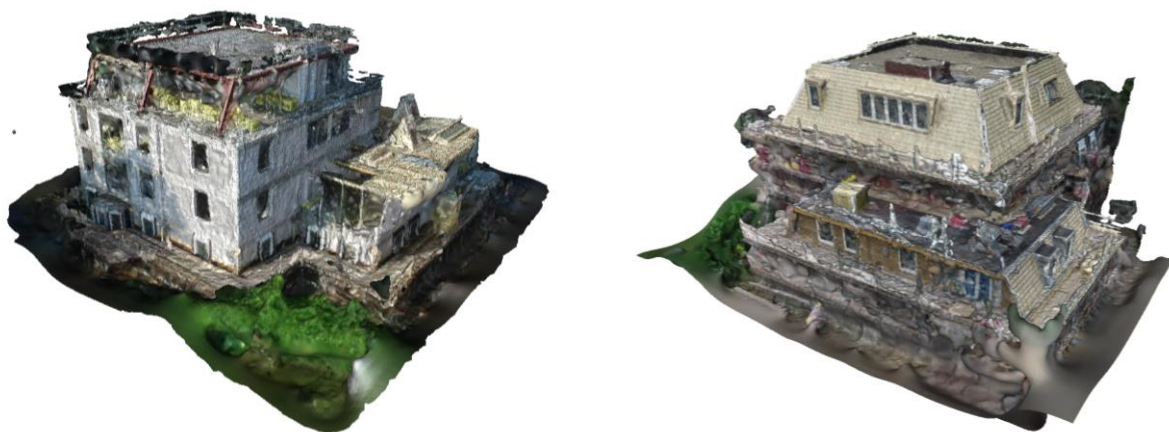


Figure 4-13 3D Reconstruction of week 27 (left) and 30 (right) with COLMAP.

4.3 Laser Demonstrator (OPC & NUIG)

The precast concrete structure used in testing the 3D Leica scanner consisted of a foundation, 4 columns, 2 beams and 2 roof slabs. At a design stage, a BIM model of the structure (Figure 4-14 left) was developed. The BIM was then used to manufacture the precast element in Oran Pre-Cast manufacturing plant. Once manufactured, the precast concrete structure was erected in the yard of the manufacturing plant. During the 3D scanning procedure, the technician distributed 4 reference points in the surrounding area. The reference points allowed for later post processing of data and linking various scans together. Suitable locations for the scans were based on: visual access to reference points, maximising the obtained visual information, extracting key features and minimising the number of scans.

During this demonstration, 10 scans were carried out (with an additional scan+photo routine). The foundation on its own was scanned twice. Next, two scans were carried out once the four columns were erected. Additional 3 scans were completed with the beams installed. Finally, when the roof slabs were positioned (and the structure was completed), the last 3 scans were carried out. An additional scan + photo was performed to allow for a colour layer that could be included at the later stage.

During the test, the scanner acquired 1,000,000 points per second and each scan took 60 seconds. The resolution of the scans was approximately 1 point / every 2 mm and the accuracy was ± 1.5 mm. Once the scans were completed, a Leica software (cyclone[®]) was required to process the data (link together the different scans) and export to a required data format.

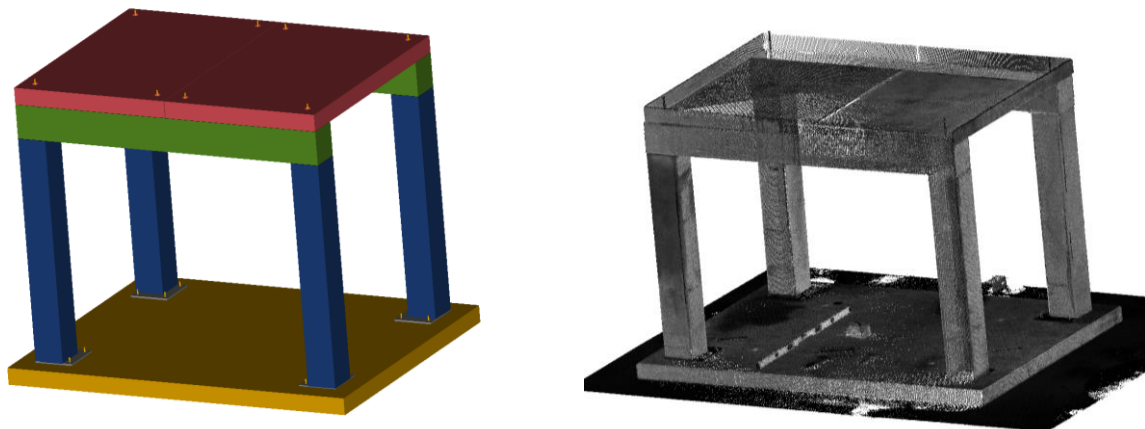


Figure 4-14 BIM and 3D reconstruction.

4.4 Eurecat Dataset

This dataset is part of the European project <http://www.fp7-icarus.eu>, it was taking in collaboration between Eurecat and ASL group from ETH, and used as starting point in the early stages of Built2Spec project. The BIM and point cloud reconstruction can be seen in Figure 4-15.

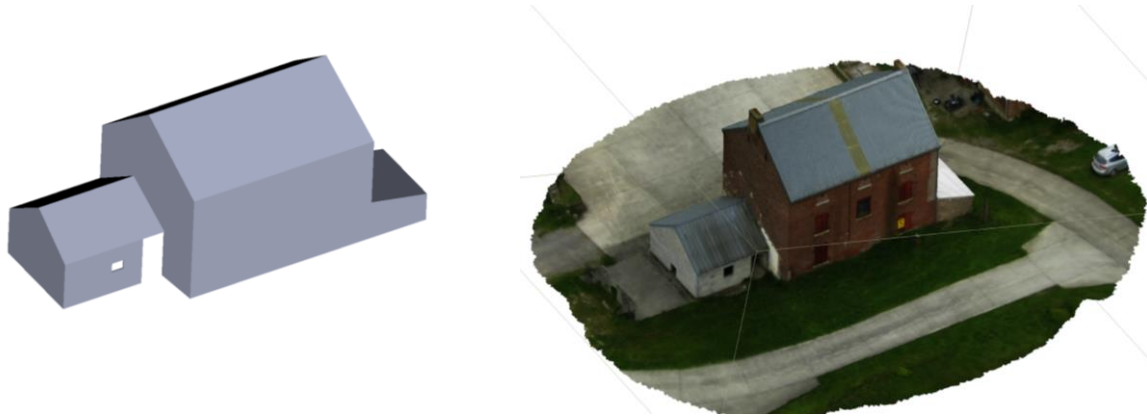


Figure 4-15 BIM and 3D reconstruction.

4.5 Acqua & Sapone Dataset (R2M & DE5)

“Acqua & Sapone” is a commercial building built by DE5 and located in San Salvo city (CH) – Southern of Italy. It is 1 floor concrete precast building about 1000 sqm (Figure 4-16).

Given the fact that the geometry of this building is quite simple, it has been chosen as demo site to develop, test and verify the workflows algorithm/software and procedures developed within T4.3 and T4.4 activities, with a focus:

- to compare a procedure to achieve high quality 3D models from drones;
- to compare the situation on site (“as built”) with the designed (“as designed”). More details will be presented in D4.4.



Figure 4-16 Acqua & Sapone building. During construction (above); finished (below).

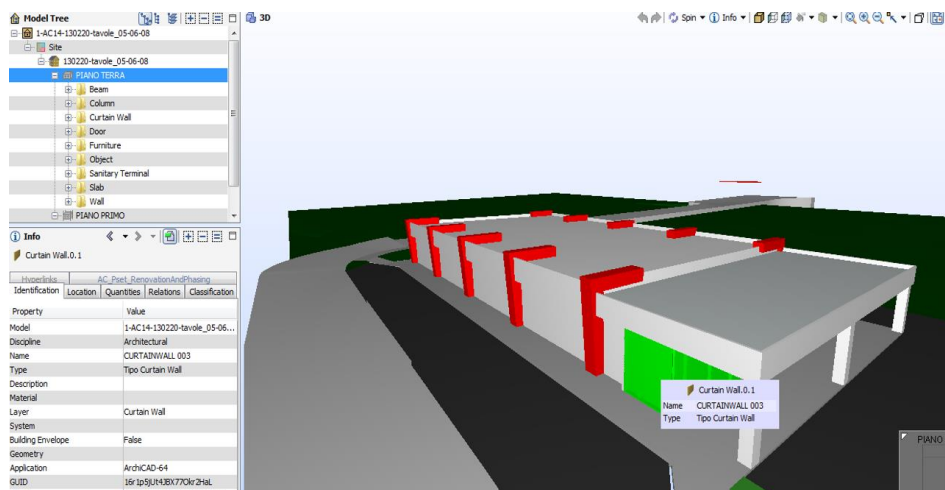




Figure 4-17 BIM and 3D reconstruction from drone images acquisition.

5. INTEGRATION WITH VCMP AND BIMSERVER

Connecting the online BIM tools to the other online parts of Built2spec is a great challenge. Recent developments in IT architecture tend towards decoupling of features. This so called ‘*microservices*’ movement is very promising. Microservices is a software architecture style in which complex applications are composed of small, independent processes communicating with each other using data-agnostic APIs. These services are small, highly decoupled and focus on doing a small task, facilitating a modular approach to system-building.

The core elements of Built2Spec are the current VRM Platform system, the BIMserver and the Point Cloud service. The focus of the Built2Spec development should be on interaction between these main components of the VCMP. The simulation tools all interact with one or more of these components.

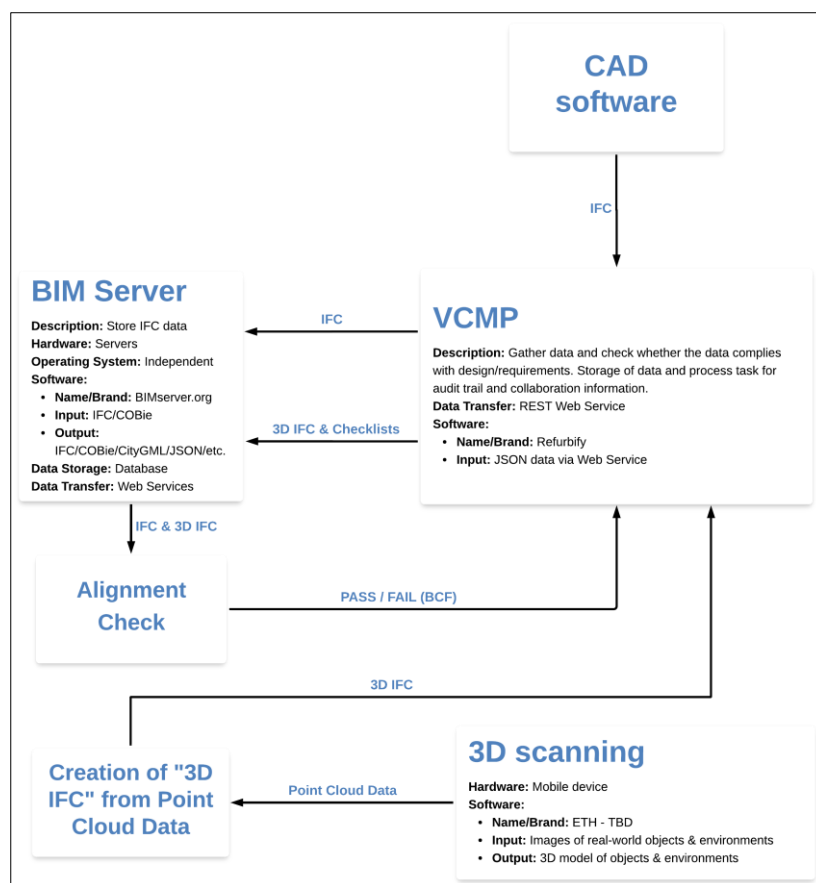


Figure 5-1 The dataflow of the VCMP and BIM/BCF.

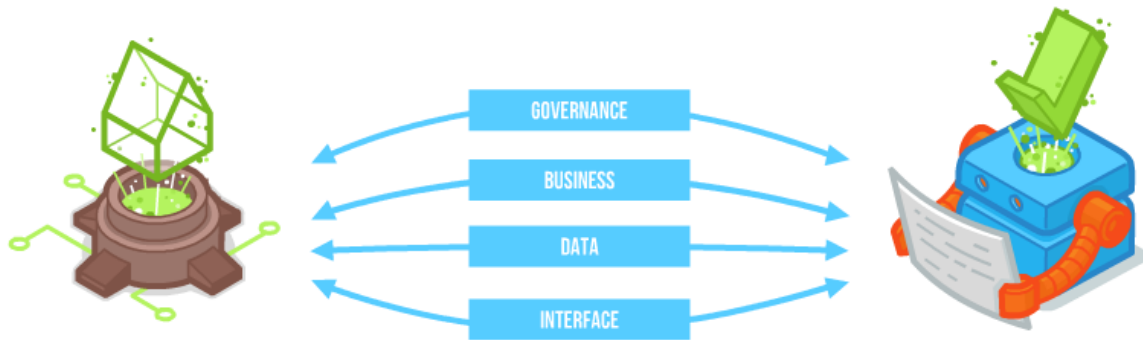
In this use-case, a user walks on site while making a point cloud of the actual build situation. The point cloud is being compared with the original design in BIM. Differences between the actual constructed object, and the designed BIM object are being highlighted and processed in VCMP.

Within Built2Spec there are several different technologies, tools and systems that have to be connected. The concept of “BIM Bots” harnesses the concept of microservices to the BIM and building industry.

A bot is an (online) system that triggers on an event. It performs a task with some form of “intelligence”. A BIM Bot can perform an analysis or simulation on your BIM data, it can enrich

the model by adding detailed objects, splitting it for specific use, or perform any other task on your data.

A user can configure a BIM Bot to run every time a certain event has taken place. By configuring multiple bots, a *'loosely coupled'* network of bots is being created for that user. The data are being send between the bots, every time enriching the knowledge and/or data.



To link different Bots to each other, there need to be agreements on four layers: Governance, Business, Data and Technical Interfaces

The several tools and systems in Built2Spec all have a unique background, interface, governance, data structure and business model. The BIM Bots concept is an ideal concept to integrate the several tools and systems, maintaining the respect to their history and background. An example of the Built2Spec BIM Bots setup for checking of the design BIM against the actual situation represented in a point cloud.

5.1 Integration with VCMP

The following description details the communication between VCMP and the 3D Reconstruction modules. An overview of this communication can be found in the workflow Figure 5-2 which breaks the communication down into the following 5 steps:

1. In **step 1**, the user is able to load a single or a series of 3D models into the VCMP system. In a previous step these 3D models can be generated by fusing 3D depth maps which can be obtained from depth sensors, or from stereo matching of images. In addition to the 3D model an acquisition date (time stamp), the construction stage of the building and the building ID is added.
2. In **step 2**, the user is able to load a single or series of BIM models (generated off-line) together with the building ID and the construction stage.
3. **Step 3** performs a comparison of scanned 3D models and corresponding BIM models. The comparison analyses models' differences and can be performed either spatial manner (single 3D model vs single BIM model) or in a temporal manner (series of scanned 3D models vs series of corresponding BIM models), which it will be described in D4.4. In order to compute these discrepancies a preceding visibility analysis of the BIM model is necessary to subsequently compute an alignment between scanned 3D model and BIM model is computed in a robust manner.

4. In **step 4**, the output of the comparison in step 3 are measures of spatial or temporal discrepancy (e.g. distance color map on the surface or volumetric differences measures, temporal progress plots).
5. **Step 5** is an optional interaction possibility for the case that the automatic alignment in step 3 fails. In this case, the user is able to perform a manual alignment of the two models by selecting at least four corresponding point pairs.

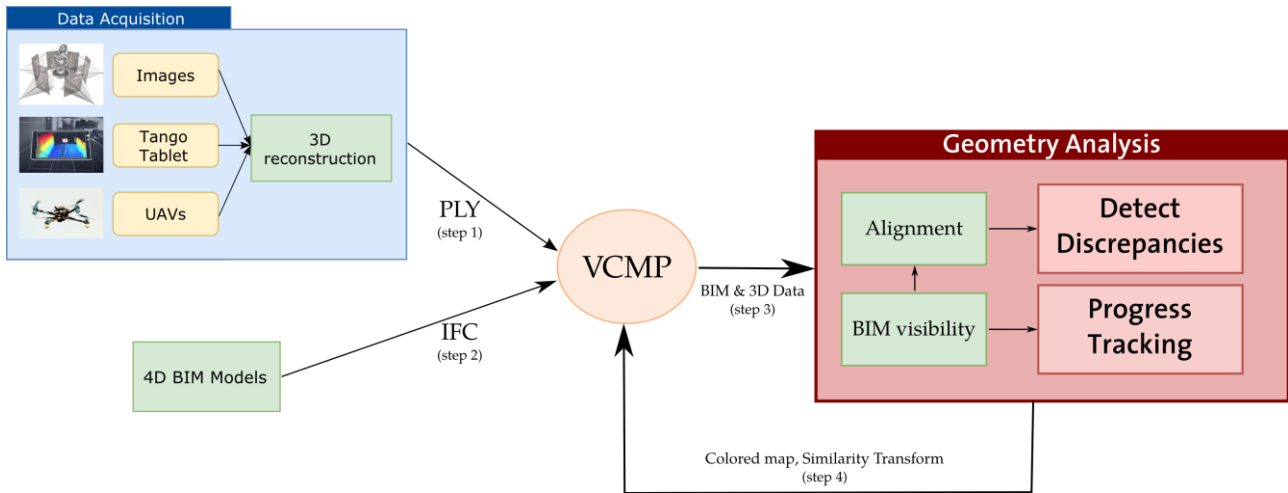


Figure 5-2 Integration with VCMP. This diagram includes the work of both Deliverables D4.3 and D4.4.

5.1.1 Inputs and Outputs

Data from BIM / information into BIM server:

- 4D BIM models (i.e. time series of 3D BIM models) from BIM (IFC format), Construction stage (in a particular time);
- (time series of) 3D reconstruction(s) (PLY format).

More details of module communication and processing steps can be found in Table 5-1.

Step	Actions	Description	Data format
1	Input to VCMP	Building 3D Scan Data (PLY), Building ID (ID code, Postal address or equivalent). <i>Source: Images / Depth Sensor</i>	PLY file. Building ID code, Dwelling No., road name, city name, country name, postcode.
2	Input to VCMP	Building BIM Model, Building ID (ID code, Postal address or equivalent). <i>Source: BIM</i>	IFC format: 4D BIM model is required (BIM models overtime)
3	Process of VCMP	Discrepancy Analysis (D4.4) 1. BIM Visibility Analysis 2. Alignment 3. Spatial- or Temporal Discrepancy Analysis	-

4	Output from VCMP	VCMP shows results of the discrepancy analysis: result of the alignment (overlay of 3D scan and BIM model), results of the discrepancy analysis (color map on the mesh, for the visible parts: true+false positives, true+false negatives of volumetric comparison).	Visibility Analysis, Alignment (Similarity transform: s (scale, 1 value), R (Rotation, 3 values) and t (translation, 3 values)). Discrepancy Analysis (Colored map representing the difference between the cloud point and BIM model).
5	User action	If the automatic alignment has <i>failed</i> , the user has to manually assist the system. This semi-automatic system requires at least 4 correspondences for the similarity transform computation.	Optionally, 4 point correspondences between the point cloud and BIM model.

Table 5-1 Step-by-step involvement of VCMP with the user or the discrepancy analysis module.

5.2 Example with BIMserver

An example use-case is explained in more details, where interaction with the BIMserver is clearer. This example is the point cloud use-case lead by ETH (Figure 5-3).

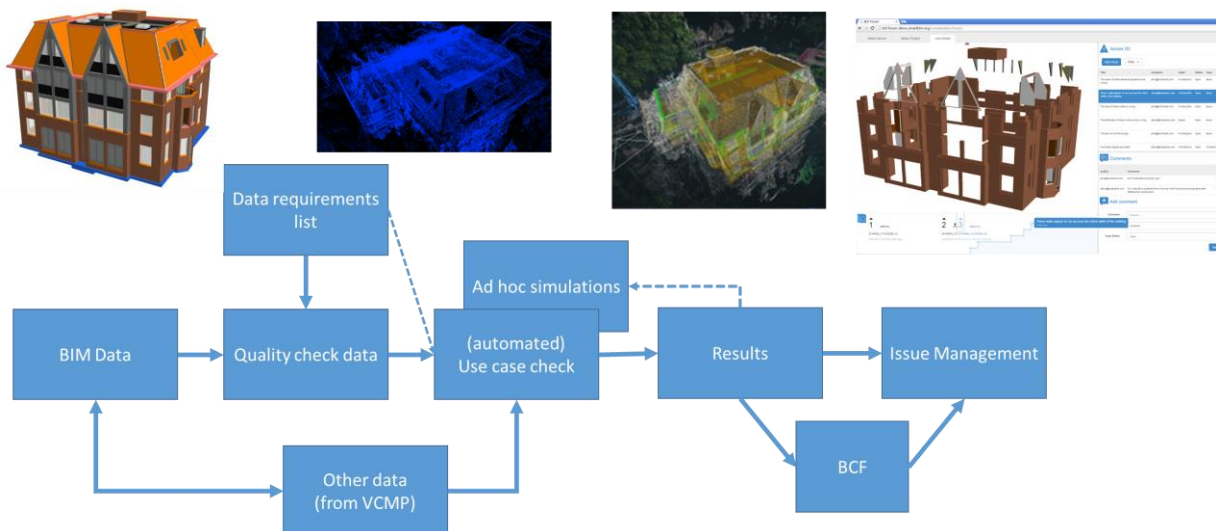


Figure 5-3 Example of point cloud use-case.

In this use-case, a user walks on site while making a point cloud of the actual build situation. The point cloud is being compared with the original design in BIM. Differences between the actual constructed object, and the designed BIM object are being highlighted and processed in VCMP.

A step by step process of the technologies used:

- A BIM model in Industry Foundation Classes (IFC) format is queried from the database (BIMserver)
- The IFC data is being checked against a predefined requirements list. The requirements per use-case, and the technology to store these requirements are being detailed in Section 5.1 of this report. In this case, the requirements to IFC are things like 'objects have to have a 3D geometry', 'the location and orientation of the elements needs to be geometrically

- exact', 'the type of elements needs to be added (walls, windows, slabs, etc.)', 'materials should be added to the objects using classification list x', etc.
- Other data, linked to the IFC data and this use case comes from the VCMP database. In this case, a list of construction margins per object is being queried. This list states the maximum allowed deviations per object type. For example, columns or walls may deviate certain threshold distance. In some cases, this additional information is being supplied by the expert user during the creation of the task.
 - Using this data, an automated check is being performed. In this case the BIM geometry is aligned with the point cloud (using an algorithm developed by ETH, which will be presented in Deliverable D4.4); objects are being recognized; deviations are being calculated; etc. This calculation is usually done in the cloud because hardware on site is not capable of performing these calculations. An internet connection is needed to start these simulations on site.
 - Finally, a result is being generated and shown to the user. Per case the result is different and it can be a number, a graph, a picture, a 3D model, etc.

In practice, communication about objects in an IFC dataset is not user friendly. This is why the 'BIM Collaboration Format' (BCF) (<http://www.buildingsmart-tech.org/specifications/bcf-releases>) is introduced. More details in Deliverable D4.5.

Depending on the result of the checks in this example, an *issue* is created. Then these issues are being processed using BCF, and can be stored in BIMserver or in VCMP, and is linked to the original IFC object.

6. ADDITIONAL WORK

Rapid BIM modelling from 3D reconstructions (ETH)

Documenting spaces with 2D media has always been difficult and designing spaces can be even more complicated. With the help of Google Tango or Microsoft HoloLens, it will be possible to capture the space in minutes (or even seconds), and automatically transforming it into an immersive digital model. Once the space is scanned, many applications can be done with it; for instance, to take measurements as it can be seen in Figure 6-1.



Figure 6-1 Take measurements in matter of seconds.

Alternately, the point cloud or mesh obtained from the 3D reconstruction could be processed to extract a CAD model or Building Information Model (BIM). Original floor plans and BIMs of a property are often not available or outdated due to past modifications. Hence, interior designers and builders often recreate building floor plans using distance measures acquired using point-to-point measuring devices. This is a tedious and time consuming task which we aim to automatize.

Furthermore, given a mesh of a 3D scanned apartment, ETH has developed a method which fully automatically computes a floor plan and a 3D BIM at interactive processing time, see Figure 6-2. This method also provides semantic output in form of plane labels (wall, ceiling, floor, door) and room labels (depicted with different colours) which can be useful for further processing such as semantic reasoning or higher-level navigation for robots. This work has been accepted to the international conference on intelligent robots and systems (IROS) 2017. A system overview can be found in Figure 6-3, and more results in Figure 6-4.

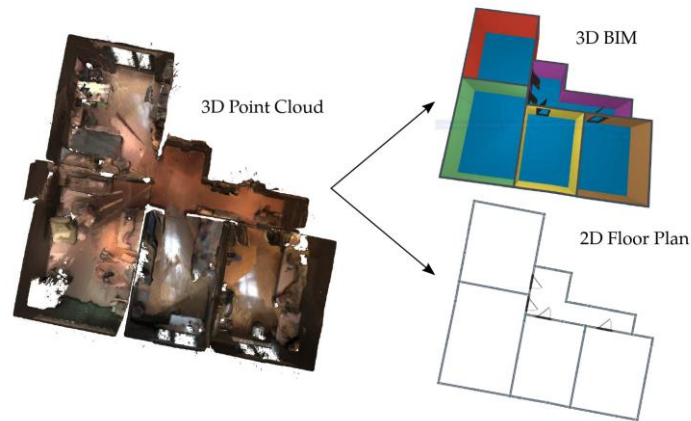


Figure 6-2 Inverse problem., from 3D Point Cloud to BIM or Floor Plan.

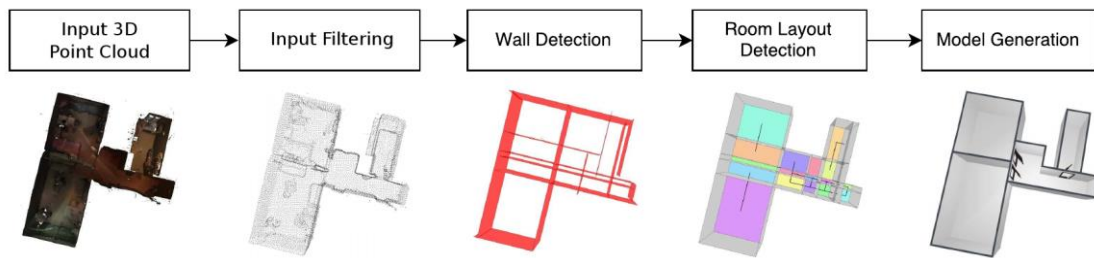


Figure 6-3 System overview with the major processing steps: 1) voxel grid filtering reduces the number of 3D points and homogenizes their density, 2) wall detection via plane candidate selection 3) room layout detection via reasoning about detected walls, and 4) model generation.

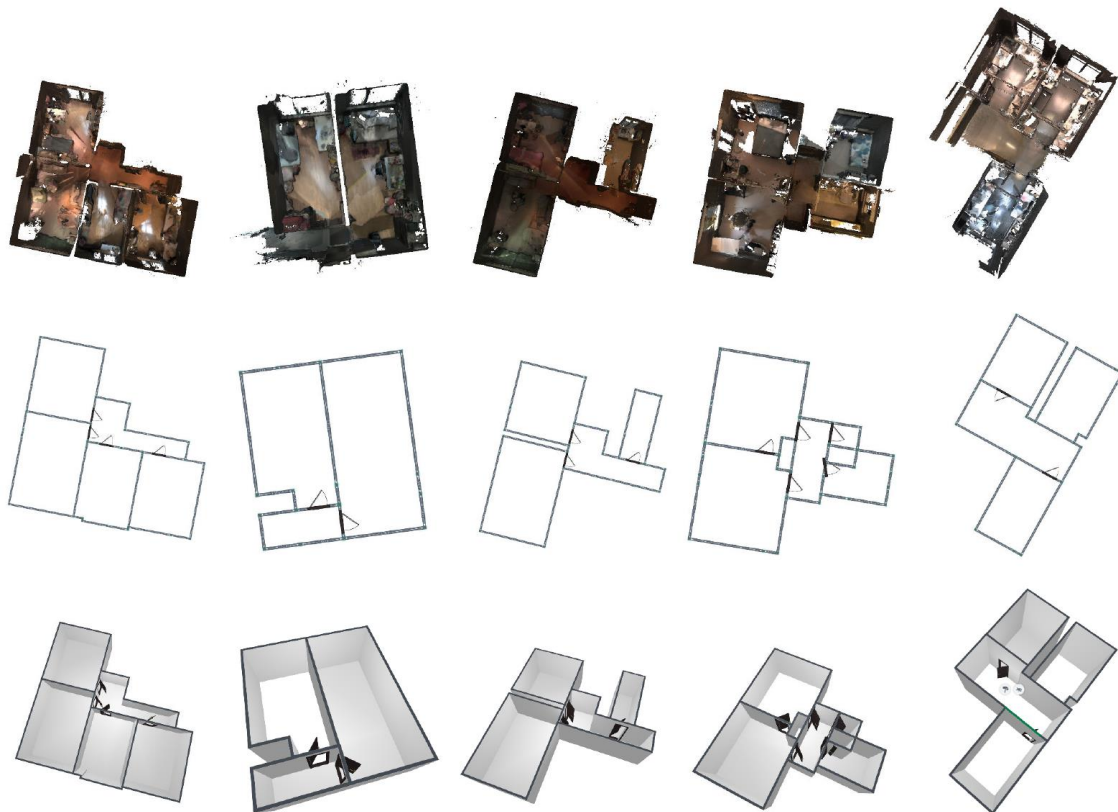


Figure 6-4 Qualitative results of our approach on five indoor datasets. Although the data is misses scanned parts and the rooms contains a lot of furniture, our method reliably computes the room layouts a variety of scenes.

7. CONCLUSIONS

For capturing 3D data, we have employed and compared five different devices, partially in combination with three different image-based 3D reconstruction techniques. In sum, the capturing devices and technologies exhibit significant differences with respect to acquisition time, post processing time and output quality, as well as hardware costs. The particular choice depends on the requirements of a particular task and the final purpose of the acquired 3D data. The quality of the captured 3D model differs considerably between the considered capturing methods. Compared to hardware costs and especially capturing time, it becomes apparent that 3D laser scanning is the most expensive and time consuming way to capture 3D data which is only worth the effort if all the other modalities do not meet the desired quality. If the scene is well textured then image-based approaches like GoPro video capturing are much faster, cheaper and are able to reach a similar quality as the laser scanner. Moreover, compared to the other technologies, the 3D laser scanning needs a special training (usually 1-day seminar held by the scanner supplier) before being able to use it. In contrast, capturing videos for image-based 3D reconstruction, or capturing scenes with the Tango or the HoloLens requires only little or no experience and is much more intuitive for the user. In contrast to HoloLens, Google Tango allows for further the storage of all input data and its offline post-processing in order to increase the quality of the reconstruction. The result of any scanning technology can be used to automatically create a BIM of the interior layout.

References to B2S Deliverables:

- **D2.2 - Building modelling and 3D automated referencing [submitted on M29 – May 2017]:** further literature review of techniques to perform 3D Reconstruction, Camera Calibration and Automatic Referencing.
- **D4.4 - Imagery Comparison Deliverable.** The output of Task 4.4 will be a document describing and evaluating on construction site data the image-based technique for identifying discrepancies between as-built and as-planned construction state [**Due on M36 – December 2017**]
- **D4.5 - Use of BIM technology for self-inspection and quality checks [to be submitted on M30 – June 2017]:** details in the integration with VCMP and BIMserver.
- **D7.2 - Real Case (non-controlled environment) Calibration and Concurrent Testing [submitted on M28 – April 2017]:** greater details about the Demo Site located in *San Salvo, Italy*. All the technologies described in **D4.3** were tested.
- **D7.3 - Long Term Testing Teams Workshop Materials [submitted on M28 – April 2017]:** further details about how to reconstruct a 3D Point Cloud starting from a survey, providing better description in how to use some of the particular software described in this deliverable **D4.3**.

Publications supported by Built2Spec project

- **Indoor Scan2BIM: Building Information Models of House Interiors.**
Srivathsan Murali, Pablo Speciale, Martin R. Oswald, Marc Pollefeys.
IEEE/RSJ International Conference on Intelligent Robots and Systems (**IROS**), 2017
- **Real-Time View Correction for Mobile Devices.**
Thomas Schöps, Martin R. Oswald, Pablo Speciale, Shuoran Yang, Marc Pollefeys.
IEEE International Symposium on Mixed and Augmented Reality (**ISMAR**), 2017
- **Consensus Maximization with Linear Matrix Inequality Constraints.**
Pablo Speciale, Danda P. Paudel, Martin R. Oswald, Till Kroeger, Luc V. Gool, and Marc Pollefeys.
Conference on Computer Vision and Pattern Recognition (**CVPR**), 2017, [PDF](#).
- **A Symmetry Prior for Convex Variational 3D Reconstruction**
Pablo Speciale, Martin R. Oswald, Andrea Cohen, and Marc Pollefeys
European Conference on Computer Vision (**ECCV**) 2016, [PDF](#).
DOI: [10.1007/978-3-319-46484-8_19](https://doi.org/10.1007/978-3-319-46484-8_19)
- **Indoor-Outdoor 3D Reconstruction Alignment**
Andrea Cohen, Johannes Schönberger, Pablo Speciale, Torsten Sattler, Jan-Michael Frahm and Marc Pollefeys
European Conference on Computer Vision (**ECCV**) 2016, [PDF](#).
DOI: [10.1007/978-3-319-46487-9_18](https://doi.org/10.1007/978-3-319-46487-9_18)

LITERATURE REFERENCES

- [1] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [2] M. Muja and D. G. Lowe, "Scalable Nearest Neighbor Algorithms for High Dimensional Data," *Pattern Anal. Mach. Intell. IEEE Trans.*, vol. 36, 2014.
- [3] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, vol. 2, no. 0. 2003.
- [4] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Commun. ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [5] F. Moreno-Noguer, V. Lepetit, and P. Fua, "Accurate Non-Iterative O(n) Solution to the PnP Problem," in *2007 IEEE 11th International Conference on Computer Vision, 2007*, pp. 1–8.
- [6] R. I. Hartley and P. Sturm, "Triangulation," *Comput. Vis. Image Underst.*, vol. 68, no. 2, pp. 146–157, Nov. 1997.
- [7] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, "Bundle Adjustment -- A Modern Synthesis," *Vis. ALGORITHMS THEORY Pract. LNCS*, pp. 298--375, 2000.
- [8] S. Lynen, T. Sattler, M. Bosse, J. Hesch, M. Pollefeys, and R. Siegwart, "Get Out of My Lab: Large-scale, Real-Time Visual-Inertial Localization," in *Robotics: Science and Systems XI*, 2015.
- [9] M. Klingensmith, I. Dryanovski, S. S. Srinivasa, and J. Xiao, "CHISEL : Real Time Large Scale 3D Reconstruction Onboard a Mobile Device using Spatially-Hashed Signed Distance Fields," *Robot. Sci. Syst.*, pp. 0–8, 2015.
- [10] M. Niessner, M. Zollhöfer, S. Izadi, and M. Stamminger, "Real-time 3D Reconstruction at Scale Using Voxel Hashing," *ACM Trans. Graph.*, vol. 32, no. 6, p. 169:1--169:11, 2013.
- [11] B. Curless and M. Levoy, "A volumetric method for building complex models from range images," *Proc. 23rd Annu. Conf. ...*, pp. 303–312, 1996.
- [12] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3D surface construction algorithm," in *Proceedings of the 14th annual conference on Computer graphics and interactive techniques - SIGGRAPH '87*, 1987, pp. 163–169.
- [13] J. L. Schonberger and J.-M. Frahm, "Structure-from-Motion Revisited," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 4104–4113.
- [14] J. L. Schönberger, E. Zheng, M. Pollefeys, and J.-M. Frahm, "Pixelwise View Selection for Unstructured Multi-View Stereo," in *European Conference on Computer Vision (ECCV)*, 2016.