MindSpaces

Art-driven adaptive outdoors and indoors design

H2020- 825079

D4.1

Analysis of multimodal signals

Dissemination level:	PU: Public
Contractual date of delivery:	Month 17, 31 June 2020
Actual date of delivery:	Friday, 03 July 2020
Work package:	WP4 Analysis of emotional, cognitive, and environmental sensing
Task(s):	T4.1: 3D-reconstruction of urban and indoors spaces; T4.2: Aesthetics and style extraction from visual content; T4.3: Textual analysis.
Type:	Report
Approval Status:	Approved
Version:	0.6
Number of pages:	114
Filename:	D4.1_Analysis_of_multimodal_signals_v0.6.docx

Abstract

D4.1 Analysis of multimodal signals [lead: U2M; due: M18; contribution: MS3]: This deliverable will present the basic techniques for 3d-reconstuction of interior and exterior spaces, aesthetics extraction and texture generation and textual analysis.

Co-funded by the European Union

History

Version	Date	Reason	Revised by
V0.1	12.05.2020	Creation of ToC	Christos Stentoumis
V0.1.1	15.06.2020	T4.3: Textual analysis	Alexander Shvets
V0.2	23.06.2020	T4.1: 3D-reconstruction of urban and indoors spaces: update	Ilias Kalisperakis
V0.3	26.06.2020	Overall update	Christos Stentoumis
V0.4	29.06.2020	T4.1: 3D-reconstruction of urban and indoors spaces: update	Ilias Kalisperakis
V0.4	30.06.2020	WP review	Christos Stentoumis
V0.5	30.06.2020	Internal review	Ayman Moghnieh
V0.6	02.07.2020	Final edit / Executive summary	Christos Stentoumis

Author list

Organization	Name	Contact Information
up2metric	Christos Stentoumis	christos@up2metric.com
up2metric	Ilias Kalisperakis	ilias@up2metric.com
up2metric	Andreas El Saer	andreas.saer@up2metric.com
up2metric	Anisa Kouka	anisa.kouka@up2metric.com
up2metric	Pavlos Bakagiannis	pavlos.bakagiannis@up2metric.com
up2metric	Lazaros Grammatikopoulos	lazaros@up2metric.com
UPF	Alexander Shvets	alexander.shvets@upf.edu
UPF	Montserrat Marimon	montserrat.marimon@upf.edu
CERTH	Petros Alvanitopoulos	palvanitopoulos@iti.gr

Executive Summary

The deliverable 4.1 aggregates the progress done in WP4 till deploying the 1st prototype of the MindSpaces platform. The components developed in WP4 are all crucial for the efficient operation of the platform.

T4.1 (3D-reconstruction of urban and indoors spaces) has to build a data collection platform and a processing pipeline that will efficiently build 3D models of real indoor and outdoor scenes for the purposes of architectural reconstruction, design and VR presentation. This scope has put the T4.1 at the very centre of the platform, as the capturing and processing processes have to be aligned with the needs of several 3D model "consumers" within the platform, e.g. architects, designers, artists, neuroscientists. This is an ongoing process, although the first loop of 3D models within the platform is completed, in order to automate some processes and uplift the capabilities of the MindSpaces platform, but also adopt the feedback recorded from the users. Difficulties till now included integration issues, and irregularities in the data capturing process, due to legislation issues in PUC1, which affected the whole processing chain. The latest has been tackled by enhancing the capacities of the data capturing platform and adapting the reconstruction approach to efficiently combine existing data, scanner, and image data.

T4.2 (Aesthetics and style extraction from visual content) provides a novel service on how to feed an artist with textures from existing artistic approaches to use them in the context of redesigning a space. Although such algorithms exist for the past two years, it is a real challenge to change aspects on a 3d model in a way that would be relevant to user needs and that it would provide exploitable results. The integration of style extraction to the MindSpaces platform has provided some interesting results, either when re-texturing a model, or when changing the texture of specific objects in the model.

T4.3 (Textual analysis) is providing the MindSpaces platform a direct input from public feedback and impressions of an outdoor space. Streams of text information, which is in abundance in web, is analysed to provide useful feedback to the architects, the artists, and the municipal authorities. Furthermore, the analysed textual data could be useful input for the artists to create new forms in art by exploiting the huge amounts of local-specific textual data floating the web. Five languages are (English, Spanish, Catalan, Greek, French) are analysed by this service. The generic textual analysis in MindSpaces is addressed as a sequence of steps: (i) morphological analysis, (ii) syntactic analysis, (iii) semantic analysis. The output of each analysis is feeding the next step, and the level of abstractness required in the semantic structures can be attained gradually. Sentiment analysis is performed in addition in parallel to the generic pipeline. Textual analysis is also connected to the KB.

The tasks are completed according to the project time plan and the general feeling is that the results are matching the initial expectation. Now that the 1st version is alive, partners in D4.1 will concentrate in deploying more elaborate solutions and in incorporating the user feedback.

Abbreviations and Acronyms

AUC	Area under the Curve
BERT	Bidirectional Encoder Representations from Transformers
CAD	Computer-Aided Design
CE	Concept Extraction
DoA	Description of Actions
DoF	Degrees of Freedom
DSA	Distant Supervision Annotation
EL	Entity Linking
GPS	Global Positioning System
GUI	Graphical User Interface
HLUR	Higher Level User Requirements
ICP	Iterative Closest Point
IMU	Inertial Measurement Unit
KB	Knowledge Base
KR	Key Results
LAS	Labelled Attachment Score
LODs	Level-Of-Details
LSTM	Long Short-Term Memory
MTT	Meaning-Text Theory
MVS	Multiple View Stereo
NER	Named Entity Recognition
NLP	Natural Language Processing
NP	Noun Phrase
OOV	Out of Vocabulary
PBR	Physically-Based Rendering
PG	Pointer-Generator network
PoS	Part of Speech
PUC	Pilot Use Case
RA	Research Activities
REST	Representational State Transfer
RO	Research Objectives
ROS	Robotics Operation System
SfM	Structure from Motion
SLAM	Simultaneous Localization and Mapping
SotA	State-of-the-Art
UAS	Unlabelled Attachment Score

UAVs	Unmanned Aerial Vehicles
UD	Universal Dependency
URI	Uniform Resource Identifier
ZCA	Zero-phase Component Analysis

Table of Contents

1 INTRODUCTION

The deliverable D4.1 is the first deliverable of WP4 and it describes the analysis done on the data which was collected in the tasks of WP3. Basically, the work described in this deliverable regards the work done in the context of WP4 towards the 1st prototype of the MindSpaces platform and the initial completion of an end-to-end operability of the platform. During the seventeen months of the MindSpaces project, data of different modalities, such as images, videos, and text, were provided from WP3 and processed in a manner that they resulted to useful building blocks for an alternative, adaptive way to redesign indoor and outdoor spaces. The state-of-the-art methodologies that were adapted by MindSpaces partners and the products, or services developed for the MindSpaces platform will be discussed in detail in the following sections.

In Section 2 the Methodology is presented via the Research Objectives, the expected results, and the timeline of the WP4.

Section 3 (Relation to) briefly presents the relation of WP4 activities to the high level user requirements as described in D7.1 (Use cases, requirements and evaluation plan).

The sections 4, 5, and 6, describe the way the visual, the style and the text data are exploited. In these sections, the authors describe the SotA and the position of this work in the current advances in the corresponding fields. The connection to the initial, raw data collected in WP3, which are described in D3.1, is also given, as well as the relation to the MindSpaces platform architecture.

Section 4 (3D-reconstruction of urban and indoors spaces) meticulously describes the methodologies applied and developed for building the high-fidelity 3D models via exploiting the data captured in T3.3, T3.4. The produced visual and geometric accurate results in the MindSpaces Pilot User Cases are presented. These results from the spatial environment that will feed the following work packages to create users' adaptive spaces.

Section 5 describes the aesthetics and style extraction from web visual content into templates to offer it to the artists and creatives, who are going to use them as a baseline to build novel artworks, or spaces. The exploitation of this services and some indicative results on the PUCs are presented.

Section 6 describes the textual analysis as a service developed for the MindSpaces platform. The goal of this task is to analyse the textual contents of the social media and of the web data crawled within in T3.5 and distil their underlying semantics so as to support the understanding of citizens' sentiments, feelings, historical inputs and needs. The related work, the current State-of-the-Art and the methodology developed for the MindSpaces platform are presented, as well as some results for the PUCs.

Section 8 concludes this deliverable.

2 METHODOLOGY

2.1 Research Objectives

The work in work package 4 has two research objectives (RO):

- RO2. 3D model extraction;
- RO3. Design, emotion content extraction and production from multimodal data;

through which it supports the innovative MindSpaces platform for the creation of user adaptive indoor and outdoor spaces.

2.1.1 RO2. 3D model extraction

The RO2 aims at the reconstruction of accurate 3D-models of urban outdoors and indoors environments. The raw data that are used for this aim are mainly visual, such as images and videos, gather either from on-site data collection campaigns, either from archived material. The generated 3D models, enriched by photographic texture, will be also integrated with semantic information (T5.3), given as an attribute in metadata form or rendered as 3D object's texture. The Research Activities (RA) that are being conducted with respect to RO2 are the following:

- RA2.1 3D model extraction from archival visual material (3D from 2D)

  The scope of this activity is to extract geometric as well as visual information about objects and scenes using multiple or single archival imagery.

- RA2.2 3D reconstruction of outdoors environments

  The goal of this activity is to extract the 3D structure of outdoor scenes from images and video captured by aerial and mobile imaging systems. The generated 3D models need to be of a high geometric accuracy and visual quality in order to serve the needs of architectural design, and, of course, the needs of Virtual Reality presentation.

- RA2.3 3D reconstruction of interior environments

  The scope of this activity is to reconstruct 3D photorealistic models of indoor environments fusing data from different 3D sensing systems. This is an open topic to research community, both for the technology required to collect and process good quality data in small amount of time, but also for the purpose of fusing 3D reconstruction process to the typical architectural processes.

Table 1 summarizes the research activities under RO2 with respect to the expected Key Results (KR) of the project.

Table 1. List of Research Activities and Key Results related to the D4.1.

Research Activities	Key Results	Deliverables
RA2.1	(**KR4**) 3D models of 2D objects and art elements	D4.1, D4.2
RA2.2	(**KR5**) Photorealistic 3D models of outdoor spaces	D4.1, D4.2
RA2.3	(**KR6**) 3D mesh with photorealistic texture of indoor spaces	D4.1, D4.2

2.1.2 RO3. Design, emotion content extraction and production from multimodal data

The RO3 aims at harvesting design and emotion appearances from different sources of visual, or textual data, which are in abundance in the internet. The key concept is to understand the information available in the web, whether this is archival information, or live information from social media, and elaborate the architect's, or artist's understanding of a space and the interconnections with society and history.

The research activities (RA) that are being conducted with respect to RO3 are the following:

- RA3.1 Aesthetics and style (design) content extraction from visual data

  This activity analyses visual content from paintings, or images of artwork, in order to extract geometrical, stylistic and other aesthetics aspects concerning specific artwork collections. It then creates metadata and templates available to the MindSpaces platform via the Design Machine to be used as a basis for inspiring architectural and artistic 3D-models.

- RA3.2 Extraction and production of content relevant to design, emotion, from text data

  This activity analyses multilingual text data from social media and online cultural and other archives, so as to extract semantically integrated information about the citizens' sentiments and views on MindSpaces PUCs, and in general, on art installations in architectural changes in a space.

Table 2 summarizes the research activities under RO3 with respect to the expected Key Results (KR) of the project.

Table 2. List of Research Activities (RO3) and Key Results related to the D4.1.

Research Activities	Key Results	Deliverables
RA3.1	(**KR7**) Image aesthetics and style	D4.1, D4.2
RA3.2	(**KR8**) Multilingual text analysis and generation	D4.1, D4.3

2.1.3 **Concept workflow**

The ambition of the MindSpaces project is better envisioned via the Concept workflow (Figure 1), which elucidates the flow of data and the structure of services and components inside the MindSpaces platform.

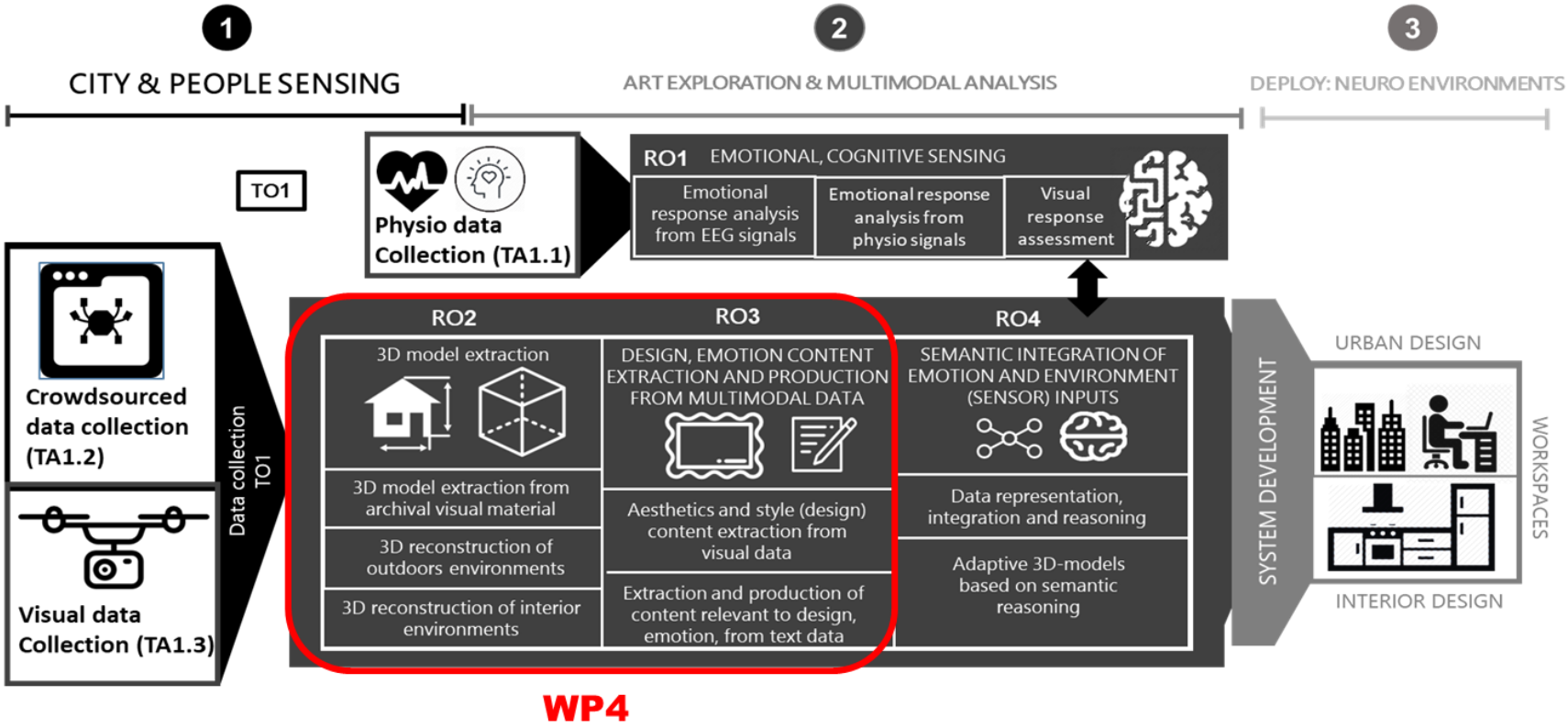


Figure 1. Research objectives and concept workflow.

2.2 Dependency Architecture

Figure 2 describes the relation of the WP4 tasks to the overall tasks' structure of the MindSpaces Project. WP4 forms the intermediate processing layer of MindSpaces platform where the collected data are pre-processed and analysed to give meaningful compounds to the MindSpaces adaptive environment. The T4.1 (3D-reconstruction of urban and indoors spaces) depends on the collection of data in T3.3 (Space sensing for 3D reconstruction) and T3.4 (Interiors sensing for 3D reconstruction), whereas T4.2 (Aesthetics and style extraction from visual content) and T4.3 (Textual analysis) depend on the web data crawled in the T3.5 (Social media and web data crawling). T4.1 (3D-reconstruction of urban and indoors spaces) and T4.2 (Aesthetics and style extraction from visual content) fuels T5.6 (Development of semantically enhanced interactive 3D spaces). The T4.3 (Textual analysis) feeds the T5.5 (Text generation).

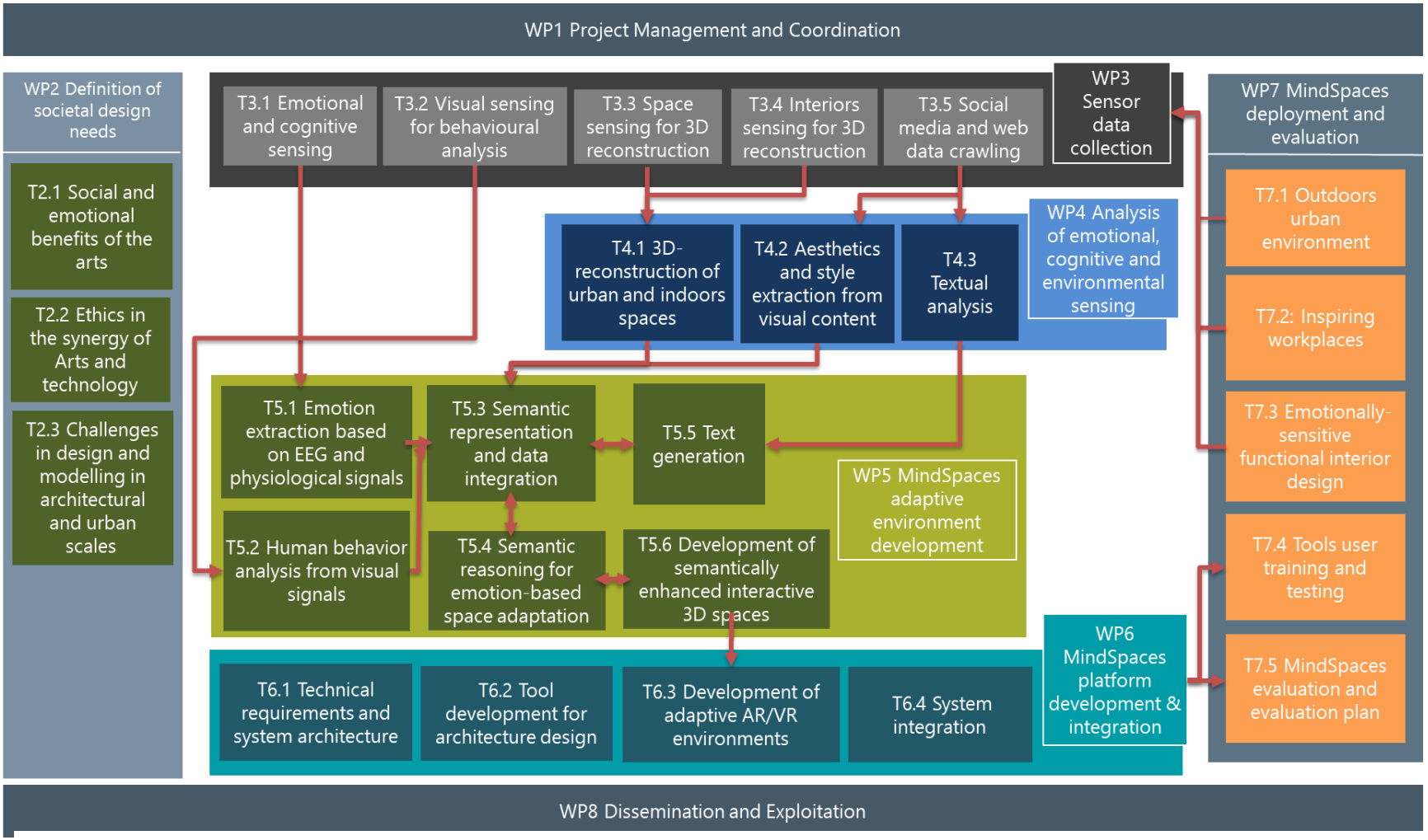
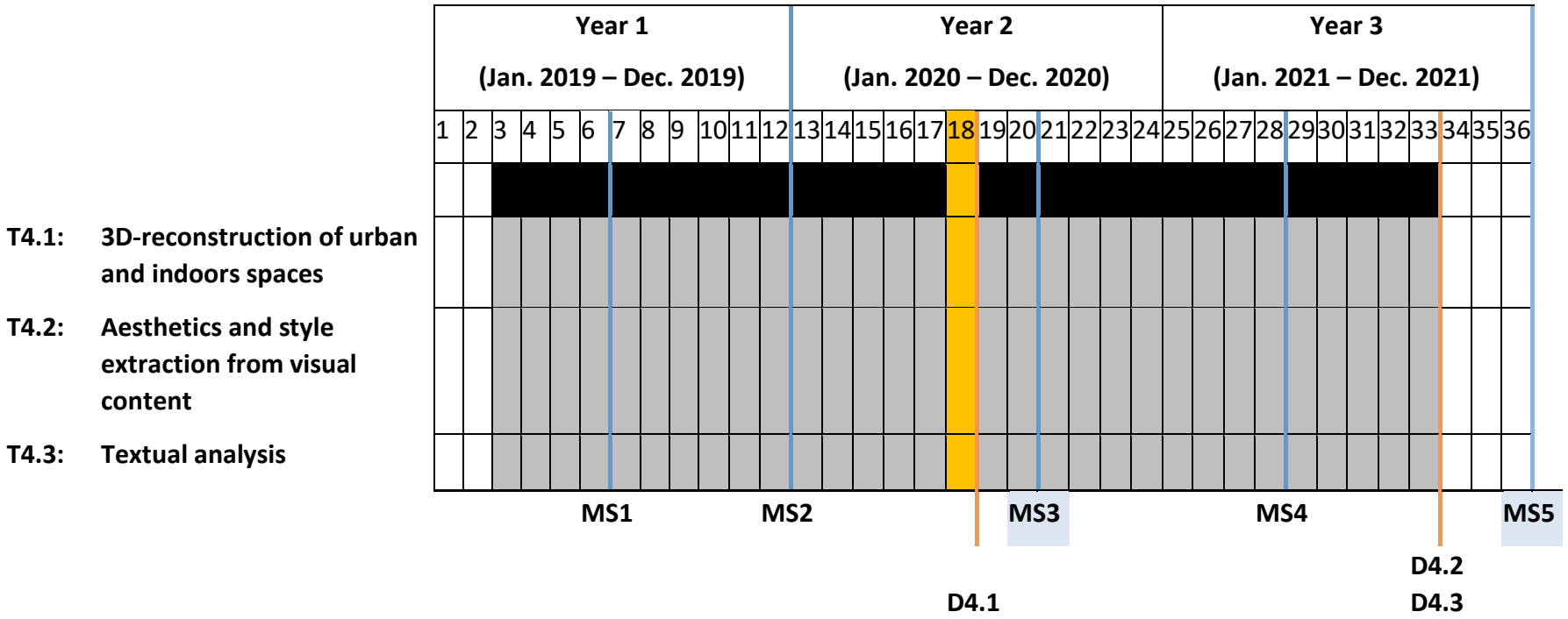


Figure 2. Dependencies architecture presenting the relation of WP4 tasks to the overall structure of tasks in the MindSpaces Project.

2.3 Timeline for Work Package 4

WP4 is at the heart of the MindSpaces platform, hence the individual tasks span across the project's whole duration (Table 3).

Table 3. Timeline for WP4.



3 RELATION TO USER REQUIREMENTS

WP4 and the tasks documented in D4.1 correlate to the Higher-Level User Requirements given in Table 4.

HLUR_2 is served through the 3D reconstruction of the indoor and outdoor spaces, whereas text analysis provides information based on the data crawled form the social media and the response of the public to a public, or private space.

HLUR_3 correlates with the textual analysis of social media data and archives on the web.

HLUR_4 is served from the T4.1 that offers the 3D model, which is a visually and geometrically exact replicate of the world outdoor or indoor space to allow the architect set the landscape of either the CAD redesign, or the VR space to get feedback from the public. The 4.2 offers a palette of artistic styles in order to offer tools to the architect, or the artist to transform space.

HLUR_5 is partially served by the textual analysis, which can provide insights based on the data trafficking on the web via social media.

HLUR_6 relates to the 3D models produced as the VR environment to get user feedback.

HLUR_9 is served via the 3D models, but also the aesthetics and style transfer service, which gives a new tool to the architects.

HLUR_11 and HLUR_12 are served via the style transfer tool.

Table 4. List of HLUR related to Data Collection tasks in WP4

	HLUR Title	HLUR Description	PUC1	PUC2	PUC3
HLUR_2	"Manipulation of spatial conditions"	Architects/Designers and artists can use spatial conditions and environmental attributes of spaces with the emotional state and behaviour of the users, to increase social interactions and communicate artistic concepts.	X	X	X
HLUR_3	"Data Analysis for understanding social needs and human values"	An artist/designer can use Data Analysis for understanding social needs and human values through social interaction with public/private spaces.	X	X	X
HLUR_4	"Adaptable spaces"	Citizens/office workers and seniors can have adaptable spaces indoor and/or outdoor depending on their needs.	X	X	X
HLUR_5	"Space use prediction"	An architect/designer can predict the potential uses for new spaces by analysing previous behavioural data.	X	X	
HLUR_6	"Intelligent projects based on feedback"	An architect/designer can produce social intelligent projects based on feedback (emotional and rational).	X	X	
HLUR_9	"Architectural design tool to form innovative ideas"	Architects and designers have a tool that can assist in formulating new, innovative architectural ideas	X	X	X
HLUR_11	"Novel textures based on the aesthetics of famous paintings"	Create novel and inspiring textures that are based on the aesthetics of famous paintings and other images of artwork that do not exist in current 3D modelling market.	X	X	X
HLUR_12	"Architects/designers aesthetics gallery"	Understand the aesthetics of design structures (i.e. interior objects, buildings, materials etc.) and provide it to architects and designers as a gallery.	X	X	X

4 3D-RECONSTRUCTION OF URBAN AND INDOORS SPACES

This section contains information about the solutions, workflows and algorithms used and developed in MindSpaces for the 3D reconstruction of urban and indoors spaces. Details about the generation of 3D models for the three Pilot Use Cases (PUC 1, 2 & 3) are also given. 3D reconstruction is performed on relevant spatial data of the environment such as raw video footage, georeferenced imagery, point clouds etc., that are collected either by industry standard surveying and photogrammetric equipment (digital cameras, 3D laser scanners, drones, etc.), or by a space sensing mobile mapping platform and a structured light scanner that are being developed in MindSpaces program. The collection of data is described in D3.1. The main objective is the production of photo-textured 3D mesh models of variant resolutions depending on the needs of adaptive design. The generated high-fidelity 3D models of urban and indoors spaces are core to the MindSpaces project since they will feed next work packages and Virtual Reality experiments to create users' adaptive spaces. The 3D models (and the raw data) will be available:

- in the platform to be accessible to artists
- in Virtual and Augmented Reality modules for adapting to Users feelings and
- in Rhino CAD for designers.

According to DoA, the 3D-reconstruction of urban and indoor spaces is involved in RO2 (3D Model extraction) and RA2.1 (3D models of 2D objects and art elements), RA2.2 (Photorealistic 3D models of outdoor spaces) and RA2.3 (3D mesh with photorealistic texture of indoor spaces) research objectives and activities of the MindSpaces project. It is expected to contribute to the following Key Results:

KR4. 3D models of 2D objects and art elements [TRL 3 to 5]: MindSpaces will use laboratory validated technologies so as to build benchmark 2D to 3D model extraction techniques and use them so as to move SotA forward and deploy robust algorithms that could function in real world settings.

KR5. Photorealistic 3D models of outdoor spaces [TRL 5 to 7]: MindSpaces will extend image-based modelling by contributing in specific tasks such as point matching among images with large pose differences.

KR6. 3D mesh with photorealistic texture of indoor spaces [TRL 4 to 7]: MindSpaces, and U2M in particular, will create a novel 3D sensing platform for indoor mapping, while simultaneously develop its core technologies, both hardware and software.

4.1 Related technologies

Both innovations in 3D computer vision and depth sensing devices and the increase of computational power have made tools and algorithms for 3D modelling of urban and indoor spaces an ongoing research field. Although commercial and industry standard methods exist, many challenges still remain when building the entire pipeline.

The development of terrestrial 3D laser scanners has allowed the fast and accurate 3D recording of complex environments. However, 3D scanning is time consuming and the captured point clouds often contain large amount of noise due to failures of reflection of the laser beams, for example on shiny, metal or glass surfaces. Additionally, 3D scanning is still considered expensive due to the high cost of using and maintaining 3D laser scanner devices. A second approach is automatic image-based modelling, i.e. photogrammetry and structure-from-motion computer vision techniques. These can capture 3D by moving a handheld video camera around an object or a space, but the respective accuracy significantly falls in cases of uniformly coloured surfaces, which is a usual case in indoor spaces (single coloured walls, floors and ceilings). Real-time depth sensors, such as Microsoft Kinect have the advantage of providing cost-effective 3D modelling. However, the obtained 3D meshes are often of low resolution and low accuracy and therefore not suitable for many application scenarios. For small scale objects, structured light 3D reconstruction methods are widely adopted, since they provide, in an automated way, high accuracy 3D models. Unfortunately, they cannot be applied for scanning objects of larger dimensions.

MindSpaces project integrates most of the above-mentioned capturing methodologies to create 3D models. These include: (a) laser scanning suitable for detailed reconstruction of large-scale objects, (b) photogrammetry and computer vision techniques that exploit visual information, (c) mobile mapping technology that combines laser scanning with Simultaneous Localization and Mapping (SLAM) and Structure from Motion (SfM) and (d) structured light 3D scanning. Tests with low-cost depth cameras are also performed to investigate the effectiveness of this approach. The integration of Unmanned Aerial Vehicles (UAVs) is also planned to facilitate the 3D reconstruction of large-scale urban environments.

As described in deliverable D3.1, several space sensing methods, tools and equipment are supported by MindSpaces platform, resulting in a scalable capturing framework. Each capturing methodology is appropriate for a specific type of surfaces (e.g., urban environments, indoor spaces, furniture, artefacts, etc.). At the same time, the combination and fusion of them, integrates the advantages/disadvantages of each method and increases the overall effectiveness of the platform.

4.2 Methodology

In this task, U2M is responsible to exploit data from different sources that are provided by the data collection services (Outdoor Spaces and Interiors Data Collection Service), in order to generate high fidelity 3D mesh models.

4.2.1 Outdoors 3D reconstruction service

In this section focus is given on the processing of relevant data for the 3D reconstruction of Urban, outdoors environments. Modifications and additional processes required for the 3D reconstruction of indoor spaces and small objects are described in Section 4.2.2 . The adopted workflow attempts the fusion of terrestrial laser scanning with image-based modelling (SfM, SLAM and stereo-matching algorithms). The processing pipeline is conceived as a separate service of the MindSpaces platform. It contains all the processes and

algorithmic solutions required for 3D reconstruction from drone and mobile mapping surveys. It can be considered as a single function, with multiple internal steps, that takes as input raw data from the “Outdoor Environments Raw Data Storage Repository” and outputs photorealistic 3D mesh models on a dedicated “Outdoor Environments 3D Models Storage” Repository in suitable format for use in Virtual and Augmented Reality apps and in CAD/CAM software such as Rhino/Grasshopper (Table 5, Figure 3).

Table 5. Main functions of Outdoor 3D Reconstruction Service

Function Name	**Description**	**Data input**	**Data output**
3D Reconstruction of outdoor environments	This function manages the generation of 3D mesh models from drone and mobile mapping surveys. It can also integrate terrestrial laser scanning and photo capture surveys.	Raw Data stored on “Outdoor Environments Raw Data Storage” data repository: *Geotagged Image Files from the drone survey* *Geotagged Image and Video sequences from the multicamera component* *3D paths and orientations from the Xsens GPS-IMU Unit and the Intel RealSense Tracking camera* *3D point clouds from the Velodyne LIDAR sensor* *3D point clouds from the Faro terrestrial 3D scanner* *Image files from the terrestrial photo capture survey*	3D mesh models with photorealistic texture

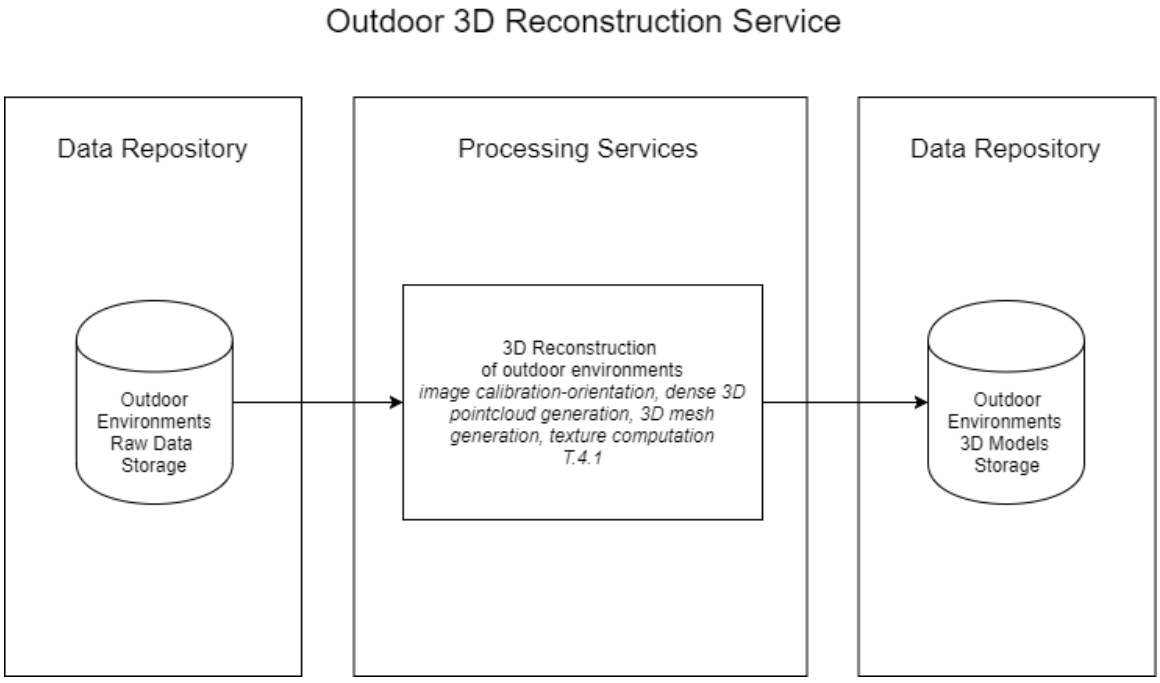


Figure 3. Logical design of Outdoor 3D Reconstruction Service

It must be noted here that there is a temporary change in the type of data used for the 3D reconstruction of Urban Spaces. In the project DoW, a commercial drone (UAV) was planned to be used for the acquisition of vertical and oblique images. Aerial imagery would be combined with street view images acquired by the MindSpaces mobile mapping platform. This approach would lead to optimal image orientations and complete modelling of the urban environments. Unfortunately, this plan is not yet realised because it was not possible to obtain a license for performing drone flight missions in the selected Pilot Use Case due to a general ban of drone usage in the specific area. A new PUC in a different Urban location is currently being organised, where the collection of both vertical and oblique image data from drone missions will be indeed realised. Instead the following data collection activities were performed with the equipment depicted in Figure 4:

i. a terrestrial photo capture mission with a DSLR camera mounted on a specifically designed telescopic carbon fibre pole, which allowed to take pictures from different heights.
ii. a 3D survey with a terrestrial 3D laser scanner, placed at multiple scan positions to achieve minimum occlusions and sufficient overlapping for the individual scans.
iii. a survey with the developed mobile mapping platform.

Activities (i) and (ii) were selected to replace the drone surveys and thus the processing of the relevant data types is added to the 3D reconstruction service. This resulted in significantly more labour work, since Structure from Motion algorithms typically underperform or even fail in such conditions and also in incomplete data, since higher elements, such as building roofs, were occluded. This said, the obtained 3D models from the post-processing are considered satisfactory and indeed suitable for specific paths in VR (hence not for flying above the roof).

Figure 4. Equipment used for the 3D reconstruction of Urban spaces. DSLR camera (1) with wide angle lens (2), a terrestrial laser scanner (3) and a mobile mapping platform (4).

The high-level workflow for the 3D reconstruction consists of the following steps:

i. Automatic image orientation and 3D reconstruction from the image data collected from drones. Here this step was performed on the imagery of the terrestrial photo capture.
ii. Registration of the individual 3D scans from the terrestrial laser scanning survey, into a unified point cloud.

iii. Automatic point cloud generation from the data of the mobile mapping platform.
iv. Mutual registration of all the obtained point clouds (from image data, the laser scanner and the mobile mapping platform).
v. Production of the final unified mesh 3D model
vi. photo-texturing of the model via the available image data

It must be noted that typically, for large scale urban spaces, steps (i) to (iii) need to be applied on thousands of images and several hundreds of point clouds. Thus, it is often preferable to divide the area of interest into subspaces which can be processed individually. If this is the case an additional step of merging the different "chunks" needs to be carried out.

Automatic image orientation and 3D reconstruction from image data (SfM)

In this step, a structure from motion scheme was implemented following an earlier approach of (Grammatikopoulos, et al. 2015) which operates on successive image scales, to facilitate the use of a large number of high-resolution images. For the implementation *AliceVision* (Jancosek and Pajdla 2011) *& Meshroom* (Moulon, Monasse and Marlet 2012) software libraries were used. Initially, stereo pairs are identified among the unordered set of images (either aerial from the drone or terrestrial from the telescopic pole). For this purpose, all images are subsampled to a low resolution; SIFT (Low 2004) and SURF (Leonardis, et al. 2006) features are extracted and a matching scheme with outlier detection (RANSAC using fundamental matrix) is applied to all possible stereo image combinations. Valid stereopairs are defined based on the number of inliers, as well as the percentage of estimated outliers after RANSAC. In case the interior orientation of the camera is unknown, an initial estimation of a common camera constant may be computed as the median of all camera constant values extracted from the fundamental matrices of all valid stereopairs (assuming that the principal point coincides with the image centre) (Sturm 2001).

Once stereopairs have been selected, SIFT and SURF features are extracted at a higher image scale and matched via RANSAC based on the five point algorithm (Nistér 2004) for the estimation of the essential matrix. Image matches are thus established across different stereopairs leading to multi-image point correspondences. A bucketing algorithm is then performed to reduce the number of tie points, without affecting their distribution on images.

For the initialization of all image exterior orientations, a stereopair is selected as reference; for every new stereopair, relative orientation is estimated from the essential matrix, tie points are reconstructed in 3D space through triangulation and a 3D similarity transformation allows inserting the current stereopair into the reference system. Local bundle adjustment solutions are held for every *N* successive images to ameliorate the exterior orientation accuracy, and a full self-calibrating bundle adjustment is performed among all available images.

Following the hierarchical scheme, new feature points are collected at successively higher image scales. Matching is restricted by the known image orientations (epipolar constraint) and by a rough 3D reconstruction of the object surface that is obtained from the tie points of previous image scale. This approach is repeated up to the full image resolution, leading to final bundle adjustment.

After image orientation, dense point clouds are generated by means of dense stereo (Hirschmüller 2005) and multi-image matching algorithms (Multiple View Stereo - MVS), followed by a triangulation in object space. These methods take advantage of the epipolar geometry derived from the exterior orientation information and determine a pixel-to-pixel correspondence between images for every image pixel, instead of distinct features only. Each pixel corresponds to a viewing ray to the object. By intersecting all viewing rays for a common, matched object point, a 3D point can be determined. By increasing the number of rays, the accuracy and reliability of the point cloud is increased. To achieve this, acquired stereo depth maps are combined with respect to their spatial resolution and their distribution in space.

Registration of the individual 3D scans from the terrestrial laser scanning survey

Typically, a 3D scanning survey involves the positioning of a terrestrial laser scanner at multiple, suitably selected stations. At each station, a 3D scan is performed that captures several millions of points around the 3D scanner. The individual point clouds from each scanning position are aligned into a uniform model by a surface matching approach on their wide overlapping areas. An initial registration is performed based on targets that are placed on the environment and features selected on the scans. A global registration of all scans is then performed through the ICP algorithm. For this workflow we utilized a commercial software that came with the laser scanner (Faro Scene[1]), together with CloudCompare[2], a widely used, open source software and the Open3D (Zhou, Park and Koltun 2018) point cloud processing library.

3D reconstruction from mobile mapping platform

In T3.3 and T3.4 of MindSpaces project a prototype mobile mapping platform was developed to assist space and interiors sensing for 3D reconstruction. The system is built on the Robotics Operation System (ROS) and utilizes multiple sensors to capture images, point clouds and 3D motion trajectories. These include synchronized cameras with wide angle lenses, a LIDAR sensor, a GPS/IMU unit and a tracking optical sensor (Figure 5). The design and the architecture of the system is modular so that it can be suitably modified to address effectively both outdoors and indoors environments. Details on the sensors involved, their integration and the initial processing algorithms are given in section 6.2 of the D3.1 deliverable.

[1] https://www.faro.com/products/construction-bim/faro-scene/

[2] https://www.danielgm.net/cc/

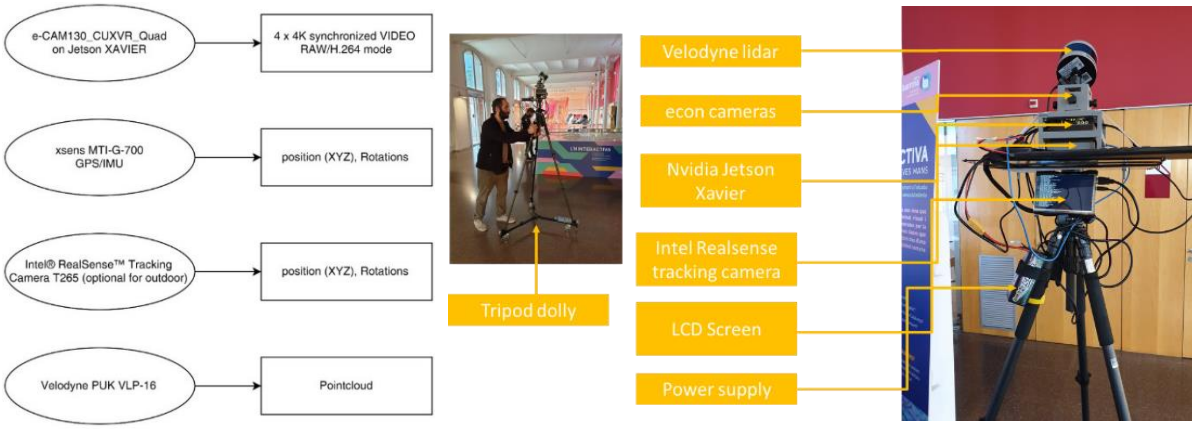


Figure 5. Sensors in the current implementation of the MindSpaces mobile mapping platform

Multiple types of data are collected from the mobile mapping platform sensors during a data collection survey. They are stored into a suitable ROS data structure, called "*bag*" file and each entry is associated with a timestamp of the moment it was captured. Thus, it is easy to retrieve all data (i.e. image frames, point clouds and 6 DoF motion trajectories) that correspond to any given time point or period and apply 3D reconstruction workflows on them. Good approximations of all sensors relative orientations are available by design, since all sensors are placed on a custom designed 3D printed case with known dimensions.

For the 3D reconstruction module of the platform existing software libraries are used. These include *ORB-SLAM2* (Mur-Artal, Montiel and Tardos 2015) and *Google Cartographer* (Hess, et al. 2016) for vision-based and LIDAR-based SLAM respectively, *AliceVision & Meshroom* (Jancosek and Pajdla 2011) and (Moulon, Monasse and Marlet 2012) for Structure-from-Motion and *Open3D* (Zhou, Park and Koltun 2018) for point cloud processing.

More specifically we use the acquired motion and rotation trajectories from the GPS/IMU sensor and the tracking camera to initialize and assist LIDAR and vision based Simultaneous Localization and Mapping solutions. These solutions provide more accurate georeferencing of the individual LIDAR point clouds and the image data from the multicamera rig.

For the LIDAR data georeferencing is crucial to obtain registered uniform point clouds of the captured space. However, the initial tests performed until this point did not succeed in providing complete point clouds of acceptable accuracy. Thus, we need to implement and evaluate additional approaches on continuous point cloud registration.

On the other hand, the initial orientation of the image frames of the multi-camera rig from the SLAM algorithms allows the selection of key poses that capture the area of interest with sufficient overlap. Then, only these key frames are used in a Structure-from-Motion workflow. Here the open source software *Meshroom* is used since it implements a self-calibration bundle adjustment solution that supports Camera Rig Calibration. This allows for optimal estimation of the four cameras interior orientation parameters along with their relative orientation. This also leads to more accurate and consistent 3D reconstruction

results. Finally, dense 3D point clouds are generated via Multi View Stereo 3D reconstruction algorithms.

Fusion of point clouds from different sources and final 3D model generation

The fusion of the acquired point clouds (from photogrammetry, laser scanning and mobile mapping platform) was performed in two steps. First, control points were extracted from the laser scanner data to fix the arbitrary scale of the photogrammetric and the mobile mapping reconstructions. Then, a registration of the datasets was performed through ICP.

Next, a fully 3D triangulation converts the 3D point clouds into 3D mesh models. Last, photorealism is achieved by computing texture for each 3D triangle via a multi-view algorithm based on a weighted blending scheme that exploits image data from an optimal subset of the available images. This ensures high visual quality for the final model, whilst eliminates seamlines and radiometric discontinuities among neighbouring images.

4.2.2 Interiors 3D reconstruction service

This service exploits all the available data from the "Interiors Data Collection Service" in order to generate and provide high fidelity 3D models of interiors and small decorative objects (T.4.1) for PUC2 and PUC3 use cases. It contains all the processes and algorithmic solutions required for 3D reconstruction from terrestrial 3D scanning and surveys with the custom-built 3D sensing platform, as well as from the structured light scanner. It can be considered as two functions, one for the interiors and one for the small objects, with multiple internal steps, that take as input raw data from the "Interiors Raw Data Storage Repository" and outputs photorealistic 3D mesh models on a dedicated "Interiors 3D Models Storage" Repository in suitable format for use in Virtual and Augmented Reality apps and in CAD/CAM software such as Rhino/Grasshopper.

Table 6. Main functions of the Interiors 3D Reconstruction Service

Function Name	Description	Data input	Data output
Interiors 3D Reconstruction	This function manages the generation of 3D mesh models from terrestrial laser scanning surveys and surveys with the custom-built 3D sensing platform. It can also integrate unordered images from DSLR cameras and RGBD image sequences from depth sensors.	Raw Data stored on "Interiors Raw Data Storage" data repository: *Individual point clouds from the terrestrial laser scanning survey* *Geotagged Image and Video sequences from the multicamera component* *3D paths and orientations from the Intel RealSense Tracking camera* *3D point clouds from the Velodyne LIDAR sensor* *RGBD image sequences from the depth sensor*	3D mesh models of interiors with photorealistic texture

Structured Light 3D Reconstruction	This function manages the generation of 3D mesh models from structured light scanning surveys	Raw Data stored on "Interiors Raw Data Storage" data repository: *Individual point clouds from the structured light scanning survey*	3D mesh models of decorative small objects with photorealistic texture

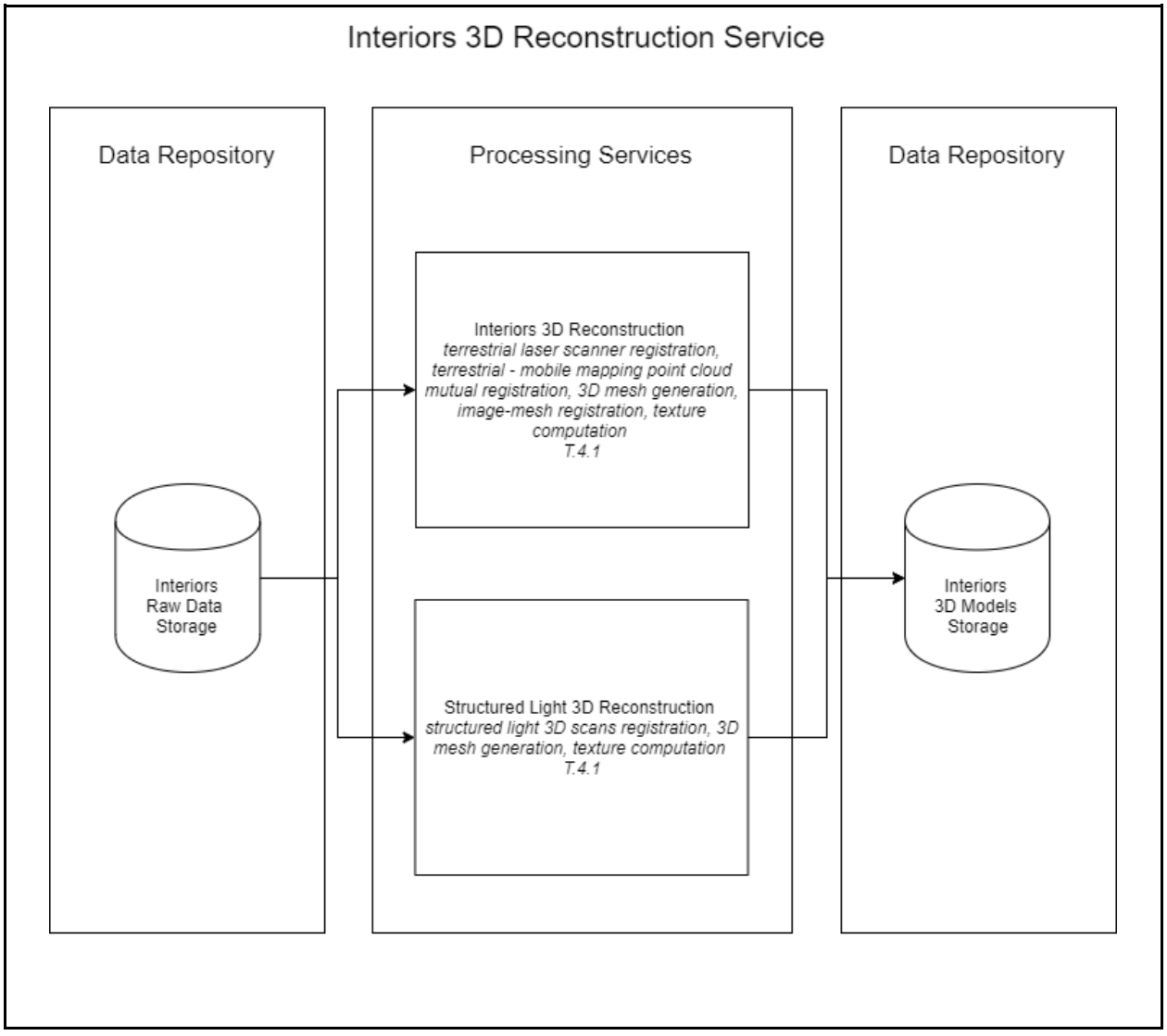


Figure 6. Logical design of the Interiors 3D Reconstruction Service

The high-level workflow for the 3D reconstruction of interiors spaces is similar to the one for the Urban Spaces with a main difference that imagery from drones is irrelevant and images from DSLR cameras are mostly used for texturing rather than modelling since, image-based 3D modelling techniques such as Structure-from-Motion often fail to provide complete 3D models in poorly textured environments. Thus, the main processing steps are:

i. a registration of the individual scans into a unified point cloud

ii. a preliminary mesh model creation from the unified point cloud
iii. an automatic image and depth sensor orientation via a Simultaneous Localization and Mapping (SLAM) scheme
iv. a mutual registration of the scanner point cloud and the image data that involves a semi-automatic step and the Iterative Closest Point algorithm (ICP)
v. the production of the final unified mesh 3D model
vi. photo-texturing of the model via the available image data.

For the 3D reconstruction of small decorative objects, a custom-built structured light scanner is employed. Structured light scanning relies on the projection of different light patterns, by means of a video projector, on 3D object surfaces, which are recorded by one or more digital cameras. Automatic pattern identification on images allows reconstructing the shape of recorded 3D objects via triangulation of the optical rays corresponding to projector and camera pixels.

In parallel to the above methodology, single image 3D reconstruction techniques that employ vanishing points and prior knowledge on objects geometry are going to be applied in paintings and 2D images, so as to build a 3D-model database that will contain interior design objects from the past.

Processing pipeline for Small Objects

In D3.1 deliverable the development of a structured light system was described in detail. It is built with off-the-shelf components and in particular with a Canon EOS 400D DSLR camera and a Mitsubishi XD600 DLP video projector (Figure 7). Details on the calibration of the system and the implemented workflow to obtain detailed 3D models were also given on sections 7.3.2 and 7.3.3 respectively.

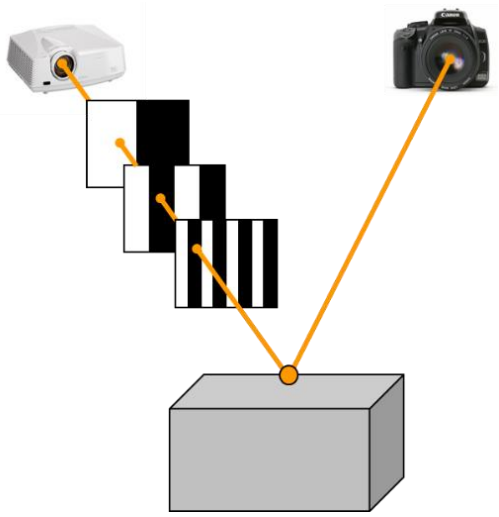


Figure 7. Overview of Structured Light system

In order to evaluate the accuracy of the developed structured light system we performed a number of scanning surveys of a small planar surface which was constructed by glass and covered by a very thin paper layer to avoid light reflectance. In particular, the planar surface was scanned with different configurations of the camera and video projector (their

distance, convergence angle and focal lengths). For each configuration, a system calibration was performed, and then the planar reference object was scanned. A plane fitting was performed to evaluate the accuracy of the reconstructed 3D mesh model. The results showed a repeatability with respect to the accuracy. The mean standard deviation of the plane fitting on all cases was of the same order (~0.030mm) and the deviations showed similar distributions (Figure 8).

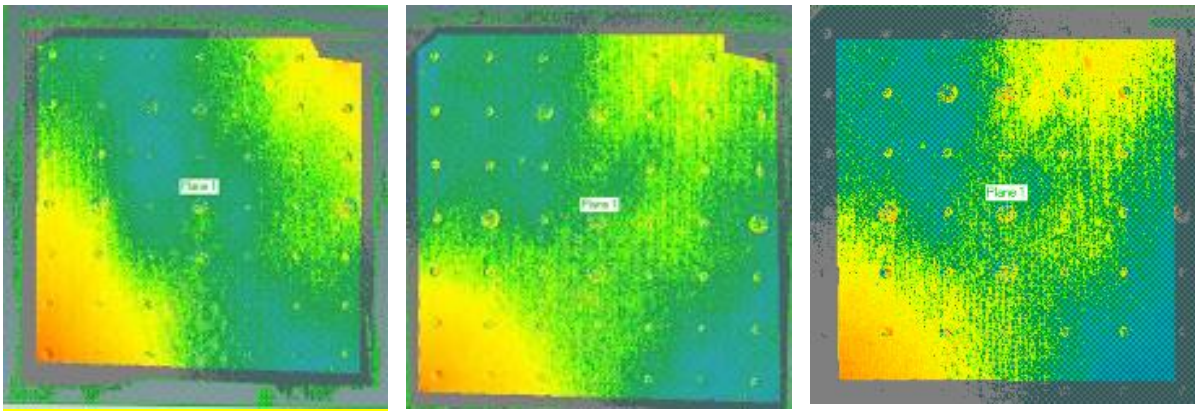


Figure 8. Deviations of a plane fitting on the reconstructed 3D model on three different scanning configurations

Then using the optimal device configuration we conducted a scan survey of a small statue of an owl, with a height of ~15cm (Figure 9). 12 separate scans were carried out from variable viewing angles to obtain a complete 3D capture of the statue (Figure 10). The individual scans were registered to a unified point cloud by means of selecting a few matching points and then via an ICP based global registration. The RMS of the registration was 0.06 mm which is an indication of the accuracy of the 3D reconstruction. Finally, a 3D mesh model was created via 3D triangulation of the unified point cloud Figure 11.

Figure 9. The owl statue which was modelled by different 3D capturing technologies.

Figure 10. Example of individual point clouds from different scan positions

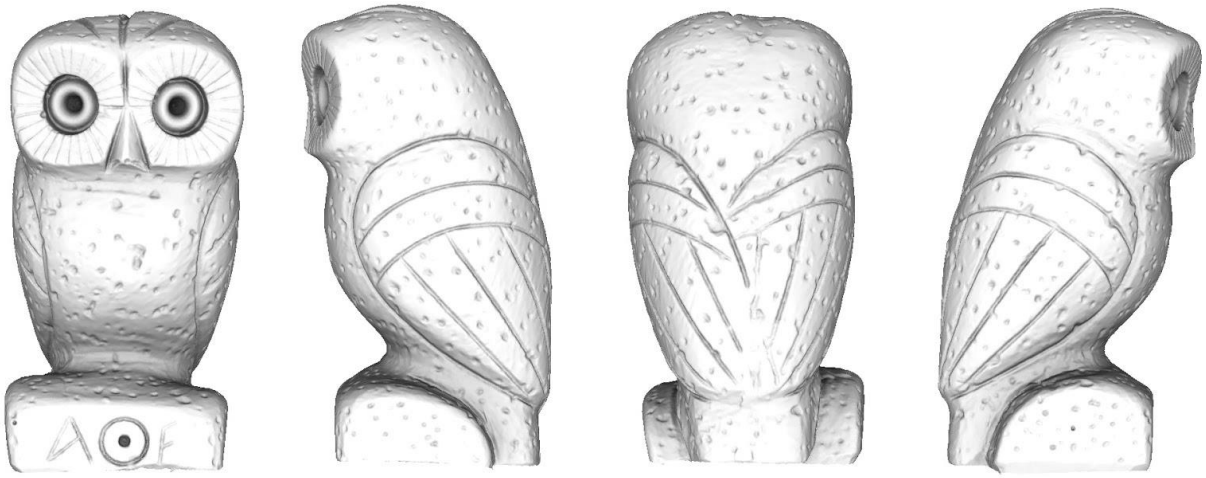


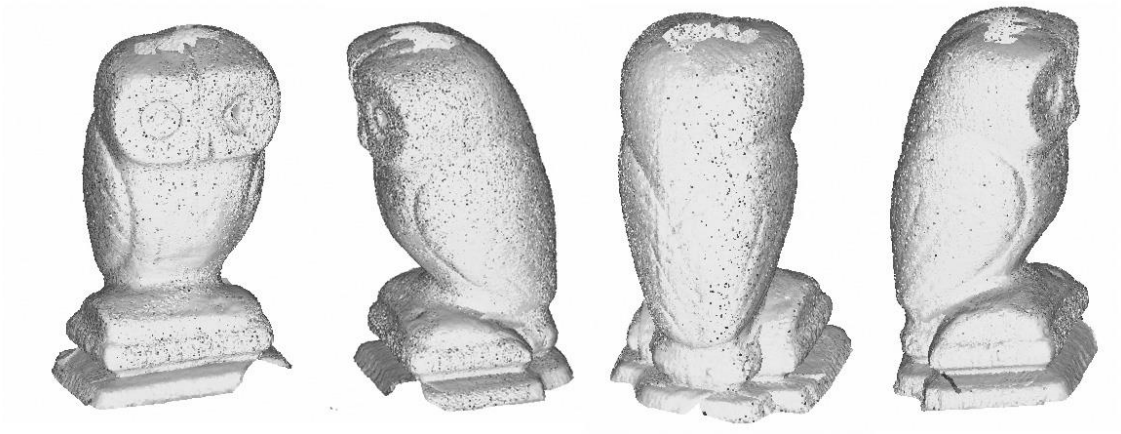
Figure 11. Final 3D mesh model reconstructed by means of structured light scanning

To further validate the effectiveness of the developed structured light system the same statue was captured by two additional 3D scanning technologies, i. image based 3D modelling by means of dense stereo matching (Figure 12) and ii. laser scanning with the Faro focus 3D scanner (Figure 13, Figure 14).

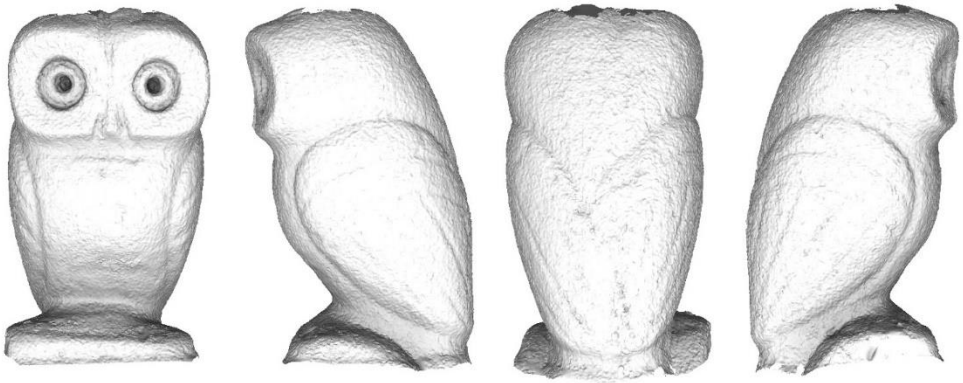
For the image-based 3D reconstruction the DSLR camera of the structured light system was used to capture the image data. To optimize the obtained 3D point cloud, two denoising algorithms were employed using the open3D library.

i. a statistical outlier removal that removes points that are further away from their neighbours compared to the average for the point cloud, and

ii. a radius outlier removal that removes points that have few neighbours in a given sphere around them.



(a)



(b)

Figure 12. (a) Example of an image-based 3D reconstruction of a small decorative object. (b) Optimization of the 3D model through standard denoising algorithms.

Denoising ameliorated the quality of the 3D model but the overall accuracy did not reach the level of the one obtained through structured light scanning.

3D scans of the statue with the Faro focus laser scanner were also carried out. As it can be seen in Figure 13 the individual point clouds are very noisy, especially on the edges. This was expected since the nominal accuracy of the specific laser scanner is ~2mm.

Figure 13. Example of individual point clouds from the Faro focus 3D laser scanner.

Similar to the image-based case, outlier removal algorithms were also applied. As shown in Figure 14 erroneous points at the edges were successfully removed. However, as expected the final 3D model was also less detailed compared to the structured light one.

Figure 14. Optimized point clouds after applying outlier detection algorithms.

4.3 3D reconstruction examples

In T3.3 and T3.4 data collection missions were carried out for three Pilot Use Cases (see sections 6.3.1, 7.4.1 and 7.4.2 of the D3.1 deliverable). In this section we present specific details of the employed workflows that were followed to obtain the 3D models for each PUC, together with the final 3D reconstruction results.

4.3.1 PUC 1 - Outdoors urban environments (Tecla Sala)

The first Pilot Use Case is the area around the cultural centre of Tecla Sala (Figure 15) which is situated in the City of L' Hospitalet, in Barcelona, Spain.

Figure 15. Tecla Sala cultural centre area overview (image from Google Earth)

The available data for this PUC consists of:

i. 6000 images from a Canon 550D DSLR camera with a 17mm wide-angle lens (Figure 16)
ii. 100 individual scans of ~3mm resolution from the Faro focus 3D laser scanner (Figure 17)
iii. 15 minutes of continuous data from the mobile mapping platform (Figure 18)

Figure 16. Sample images from the image capture survey. Images were captured at multiple heights and different camera orientations.

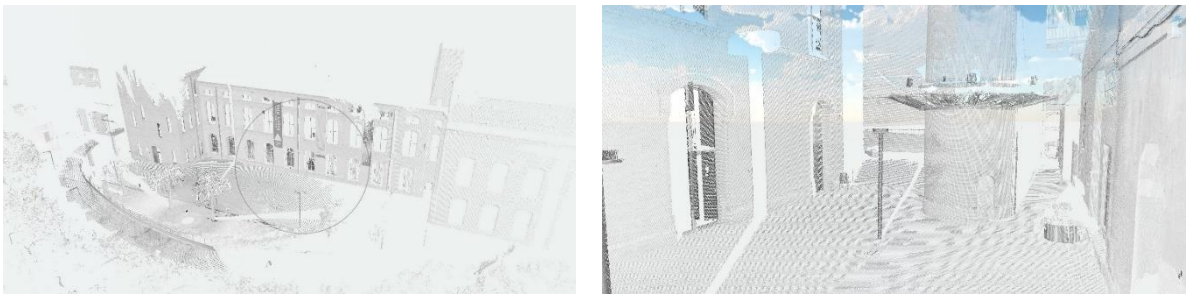


Figure 17. Sample point clouds from two different scan stations

Figure 18. Four synchronized frames from the multi rig camera of the mobile mapping platform

For the 3D reconstruction of Tecla Sala area, 3D scanning was used in combination with photogrammetry following the methodology described in Section 4.2.1 Figure 19 depicts schematically the processing steps involved.

All individual scans were registered in Faro Scene Software by means of matching targets, interest points and surfaces. A unified point cloud of 3mm resolution was generated consisting of more than 756 million points. As it can be seen in Figure 20 this included points outside the area of interest as well as outliers. Thus, the point cloud was cleaned through a combination of outlier removal and manually selecting areas of points that were out of scope. The cleaning process lead to a model of 598 million points, which was still very big for further processing.

The model was divided into seven slightly overlapping sub areas, which were processed separately. Each area was decimated, and then converted to 3D mesh via 3D triangulation. The generated triangular meshes were further simplified (Figure 21) and then merged together into a single 3D mesh model (Figure 22).

The images of the terrestrial photo capture were divided into separate chunks, one for each main wall of the Tecla Sala Cultural Centre. Each chunk was relatively oriented automatically through a structure from motion scheme on successive image scales. To assign correct scale and register the image orientations with the 3D model from the laser scanner, characteristic features on the model were measured and identified on the image set. Once sufficient control points were measured a bundle adjustment solution was performed for all images, yielding optimal estimations of image orientations (Figure 23).

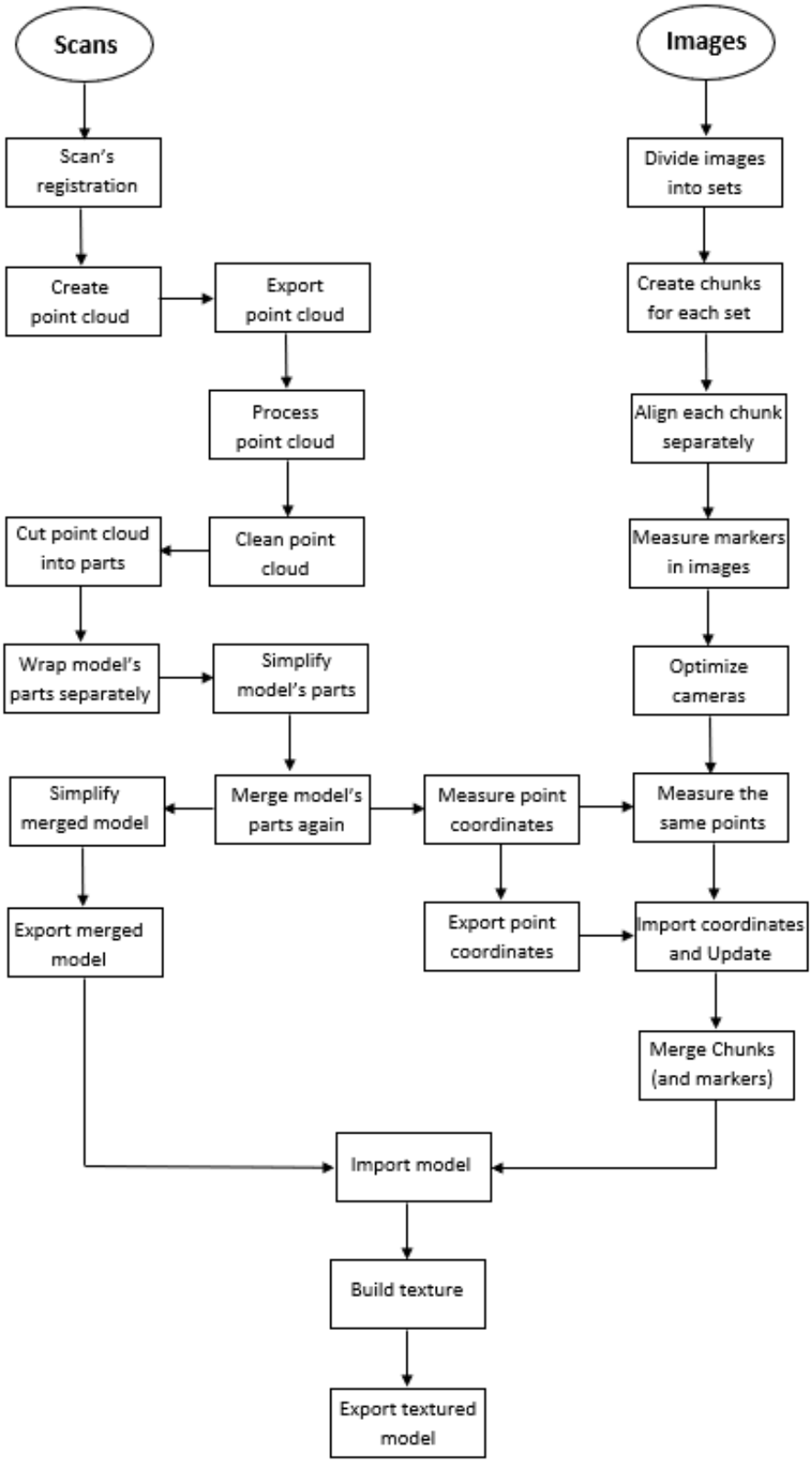


Figure 19. 3D reconstruction of urban spaces pipeline

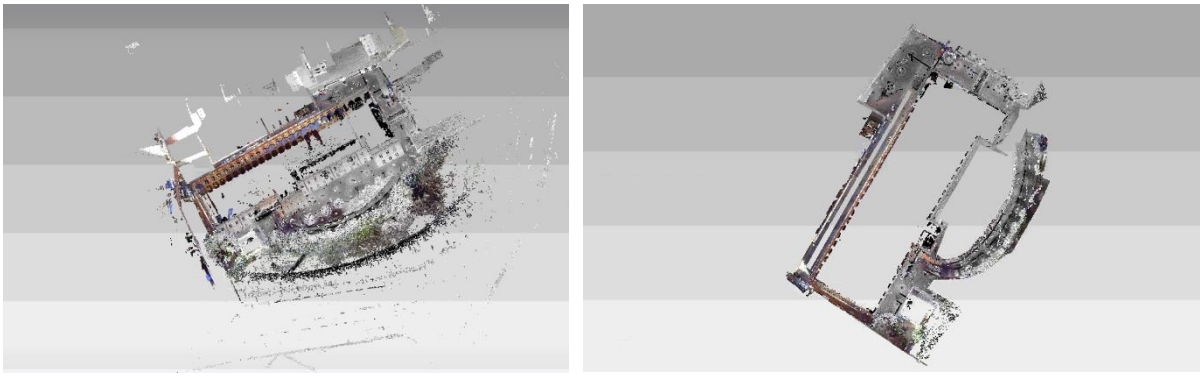


Figure 20. Original unified point cloud (~756M points) (left). Cleaned point cloud (~598M points) (right)

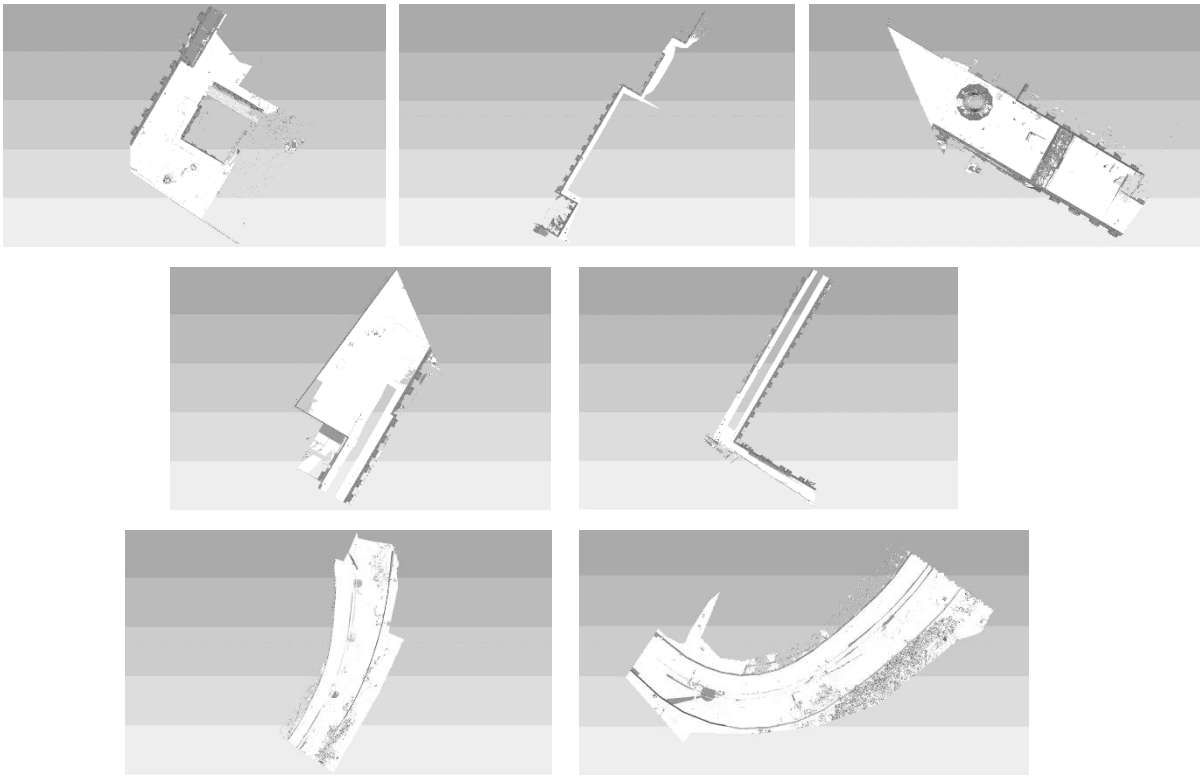


Figure 21. Simplified 3D mesh models of the divided sub areas. Each model consists of ~2M triangles

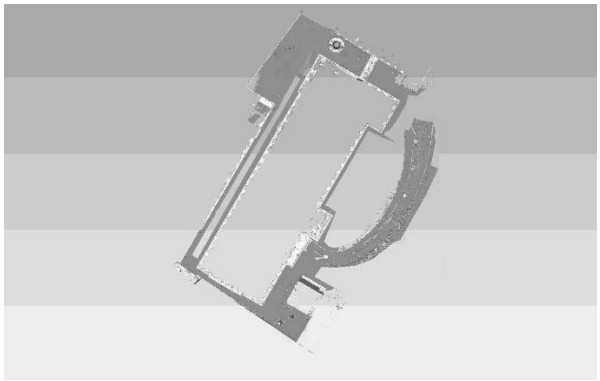


Figure 22. Merged 3D model consisting of ~14M triangles

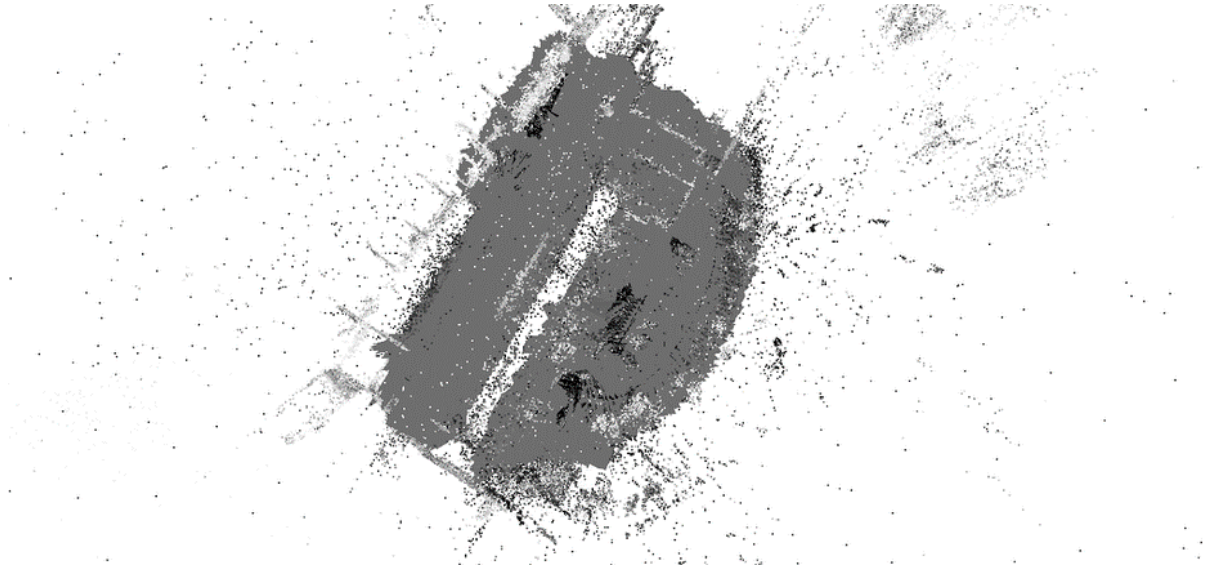
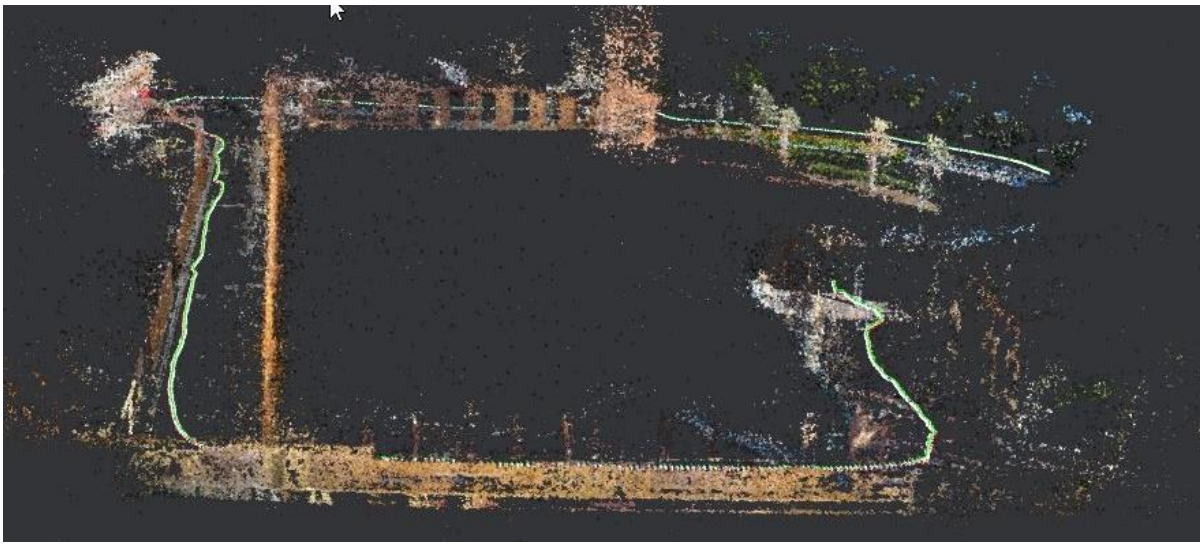


Figure 23. Final image orientations obtained through a global bundle adjustment solution

A point cloud from the mobile mapping platform data was also generated and parts of it were used to fill gaps and occlusions in the 3D model from the laser scanning survey. A vision-based SLAM solution was used to estimate the motion and rotation trajectory of the platform and then an automatically selected subset of the collected image dataset was fed to the Structure-from-Motion workflow (Figure 24, Figure 25). A dense point cloud was also computed via Multi View Stereo Dense Reconstruction.



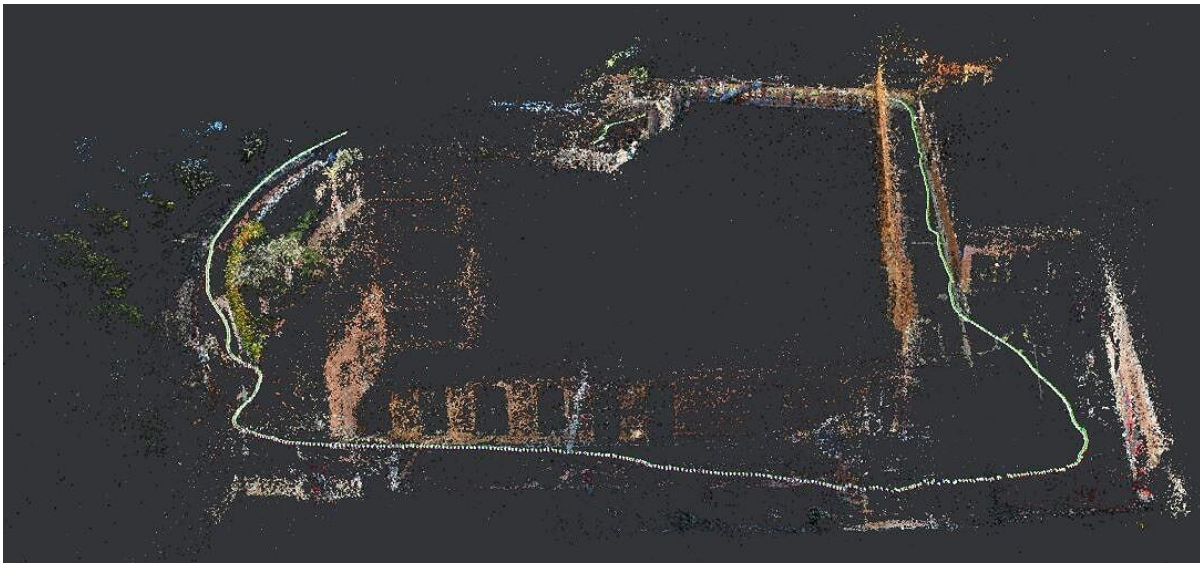


Figure 24. Estimation of mobile mapping platform trajectory by means of SLAM and SfM

Figure 25. Detail of the extracted orientations (upper) and 3D model (lower) from the mobile mapping platform data

Once all models were unified and merged into a single model, the oriented image set from the terrestrial photo capture survey was used to estimate optimal photorealistic texture to the generated 3D model, via a multi-view texture blending algorithm (Figure 26, Figure 27).

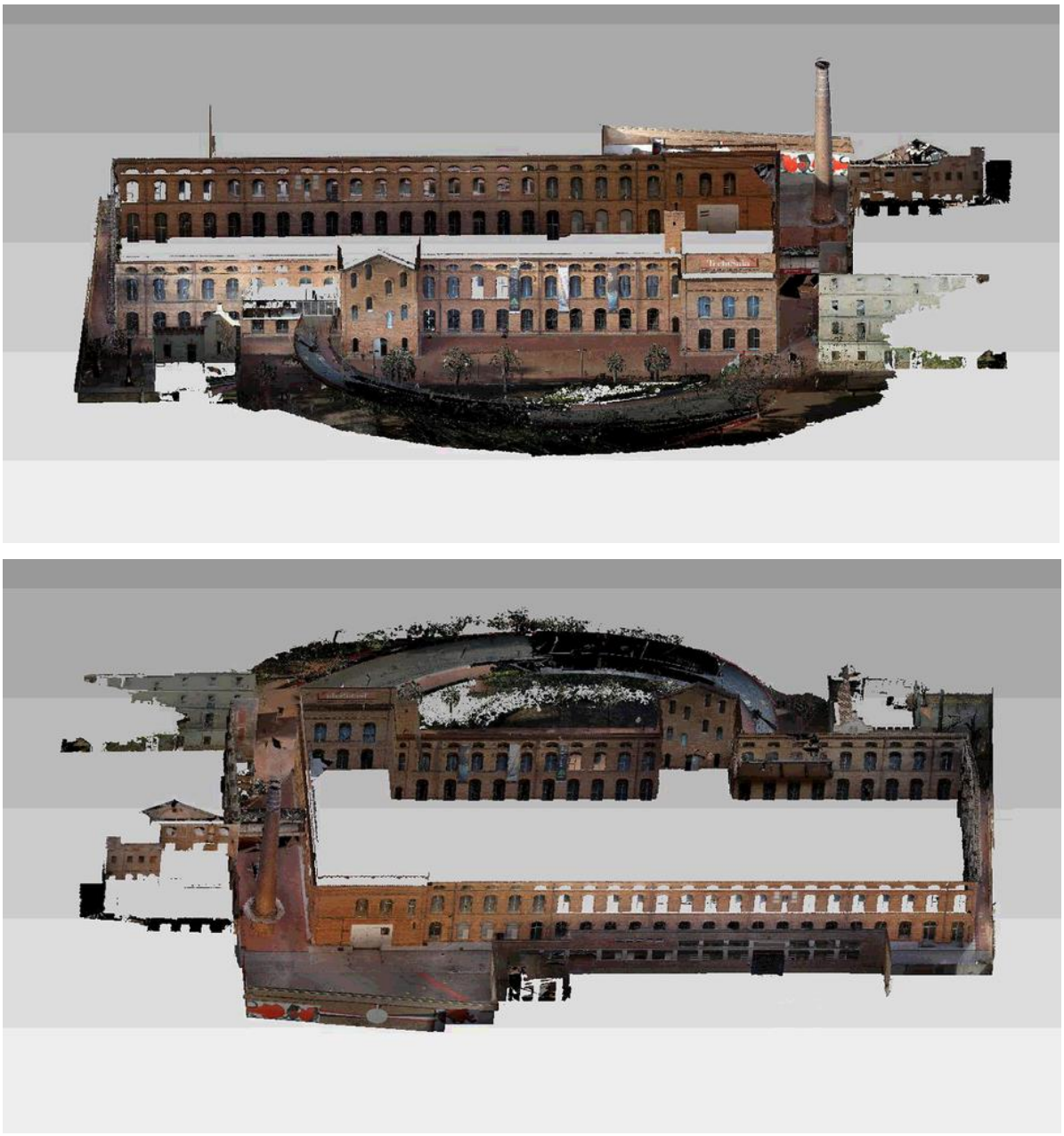


Figure 26. Final textured 3D mesh model of PUC1

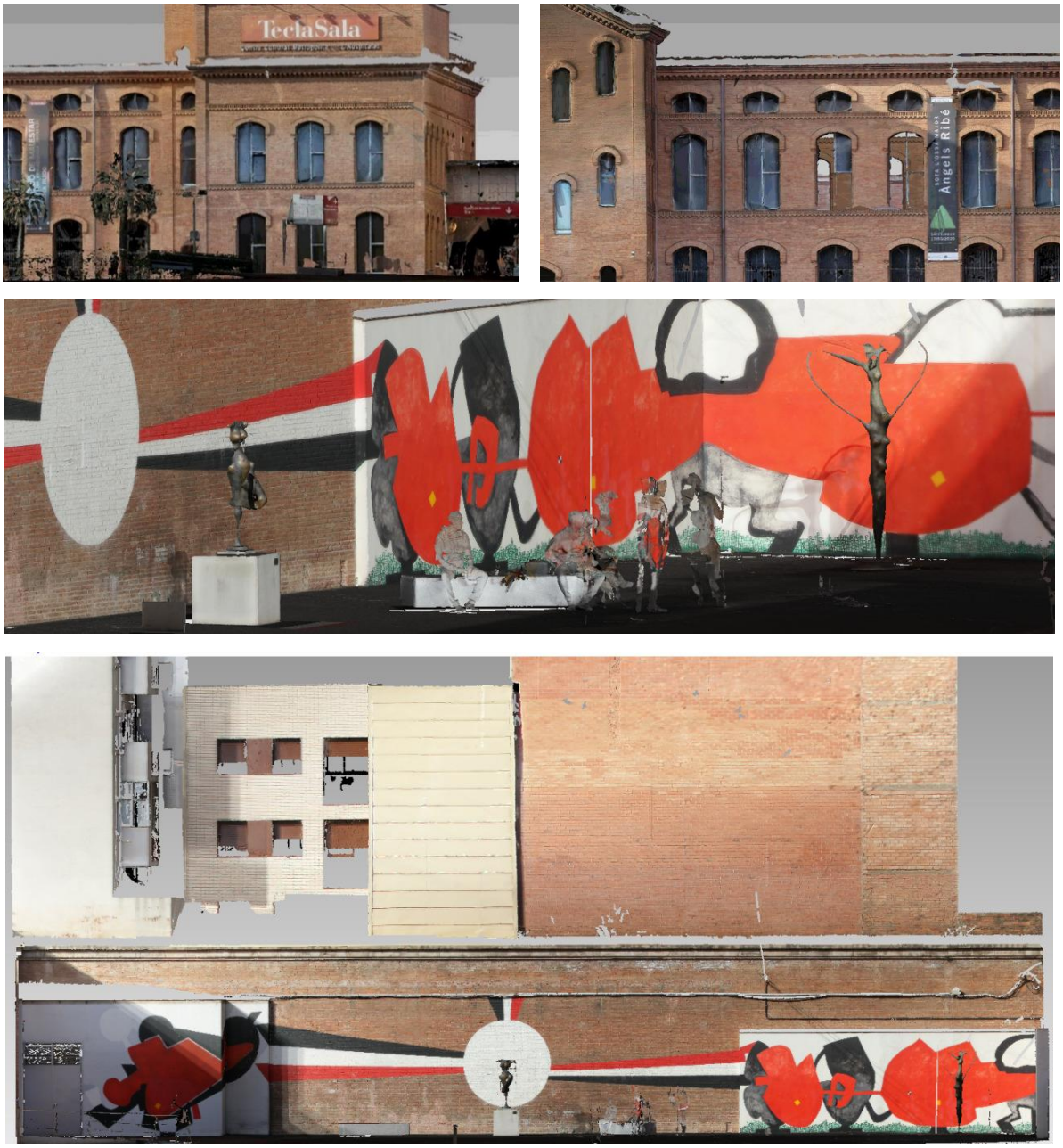


Figure 27. Details of the final textured 3D mesh model of PUC1

4.3.2 PUC 2 - Inspiring workplaces

The second Pilot Use Case was the office facilities of MindSpaces partner McNeel, which is situated in Barcelona, Spain.

The available data for this PUC consists of:

i. 1500 images from a Canon 550D DSLR camera with a 17mm wide-angle lens (Figure 29)
ii. 50 individual scans of ~3mm resolution from the Faro focus 3D laser scanner (Figure 28, Figure 17)

Figure 28. Sample point clouds from the 3D scanning survey

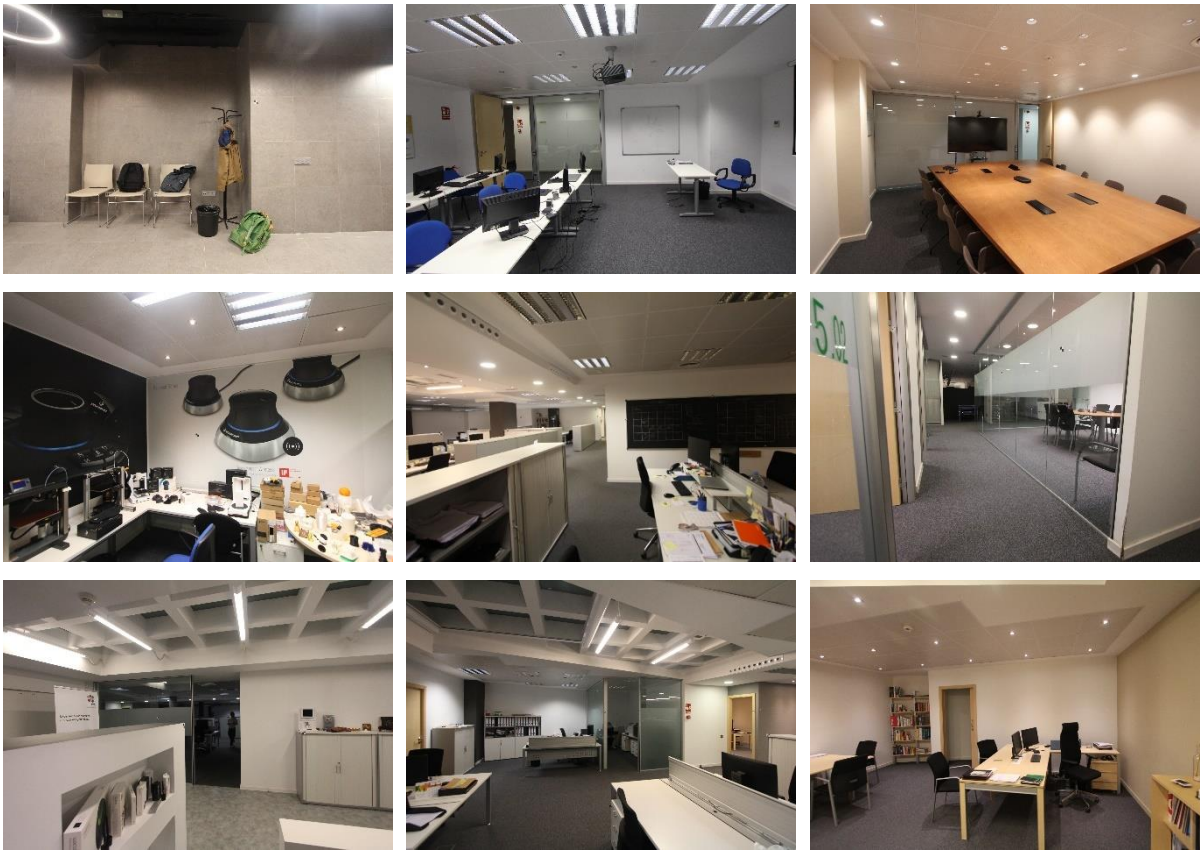


Figure 29. Sample images from the image capture survey.

For the 3D reconstruction of McNeel's offices, 3D scanning was used in combination with photos for estimating optimal photorealistic textures via a similar approach to the 3D reconstruction of Tecla Sala area. Figure 30 depicts schematically the processing steps involved.

The individual 3D scans were registered into a unified point cloud of ~3mm resolution, which was cleaned from outliers and out of scope points and divided into sub areas, one for each room. Each room was triangulated separately and a 3D mesh model per room was generated and simplified (Figure 31).

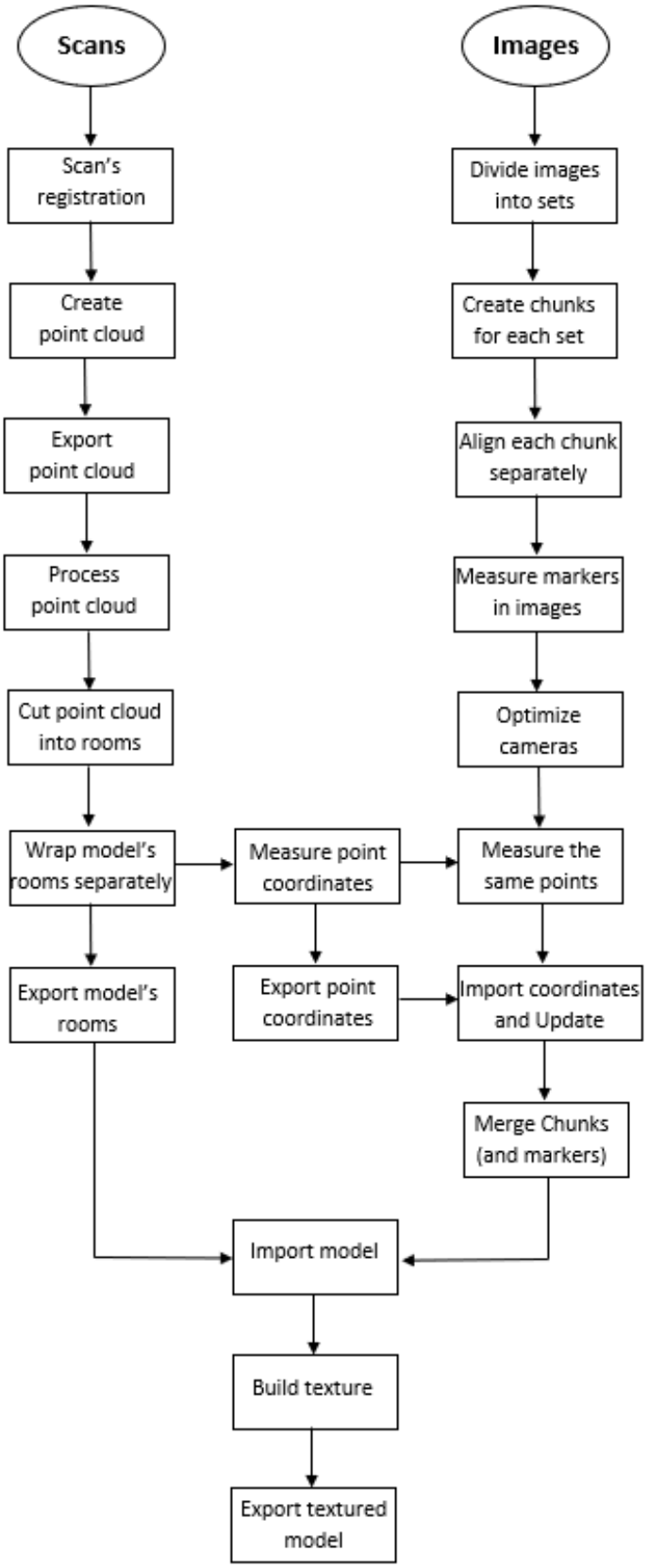


Figure 30. 3D reconstruction of interior spaces pipeline

Images were divided into chunks, again one per room and oriented automatically via SfM solutions. Control points were measured on the 3D model from the laser scanning survey and identified on the images and an optimal mutual registration of images and 3D model was achieved through a global bundle adjustment solution. Finally, texture was applied at each room from the respective images. Details of the final 3D photorealistic mesh models are presented in Figure 31 and Figure 32.

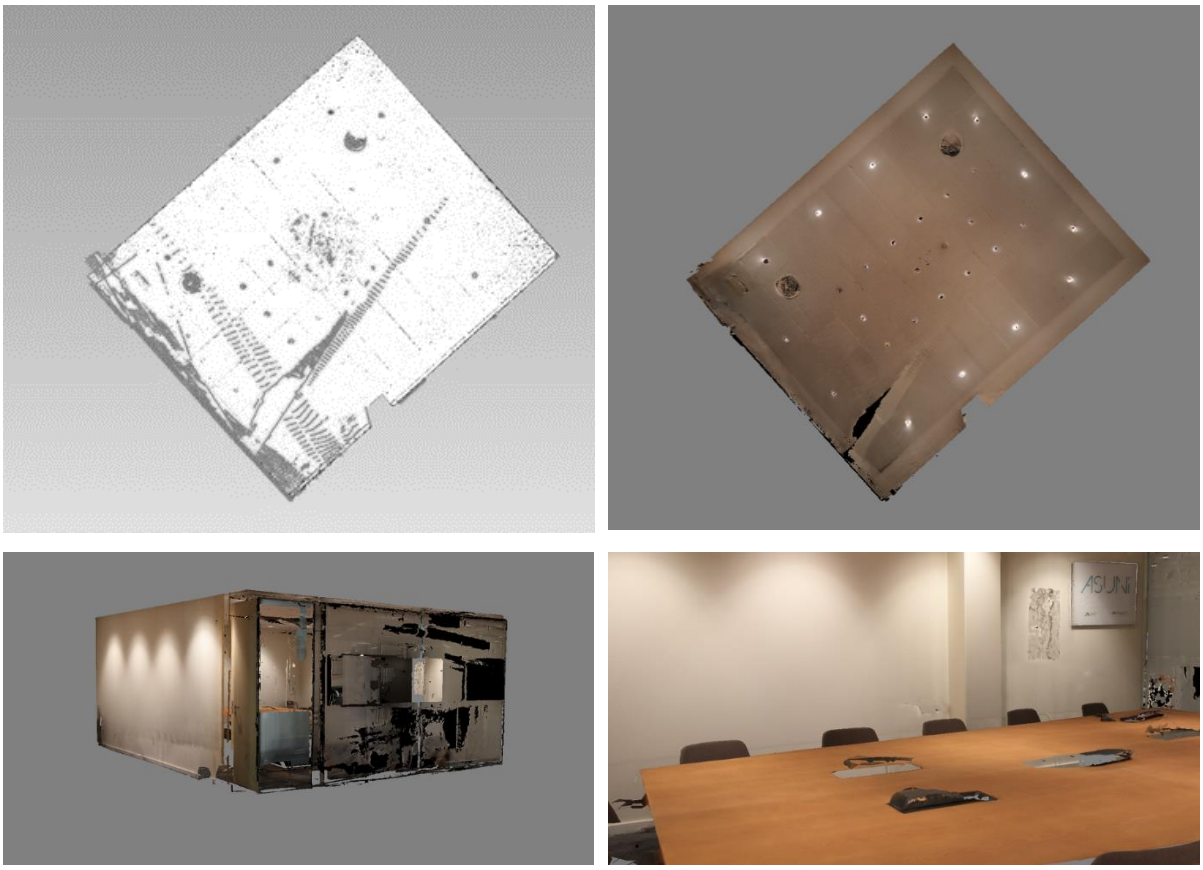


Figure 31. 3D mesh model of a room of PUC2

Figure 32. 3D mesh model of a room of PUC2

4.3.3 PUC 3 - Emotionally-sensitive functional interior design

The third Pilot Use Case was the apartment of a senior person in Paris.

The available data for this PUC consists of:

i. 20 individual scans of ~3mm resolution from the Faro focus 3D laser scanner (Figure 33)
ii. images of the apartment taken with a DSLR camera
iii. images of small decorative objects and selected furniture
iv. data from a first beta version of the mobile mapping platform using only two sensors, the Intel RealSense T265 Tracking camera for relative positioning and the Microsoft® Kinect v.2 Depth Camera to acquire depth images and point clouds

Figure 33. 3D scanning of the PUC3 residence

A textured 3D mesh model of the apartment was generated following the same procedure as in PUC2. Since the apartment was significantly smaller than the office space of PUC2, the division of the point cloud into subsets for each room was not necessary and was skipped. The final textured 3D mesh model is presented in Figure 34 and Figure 35.

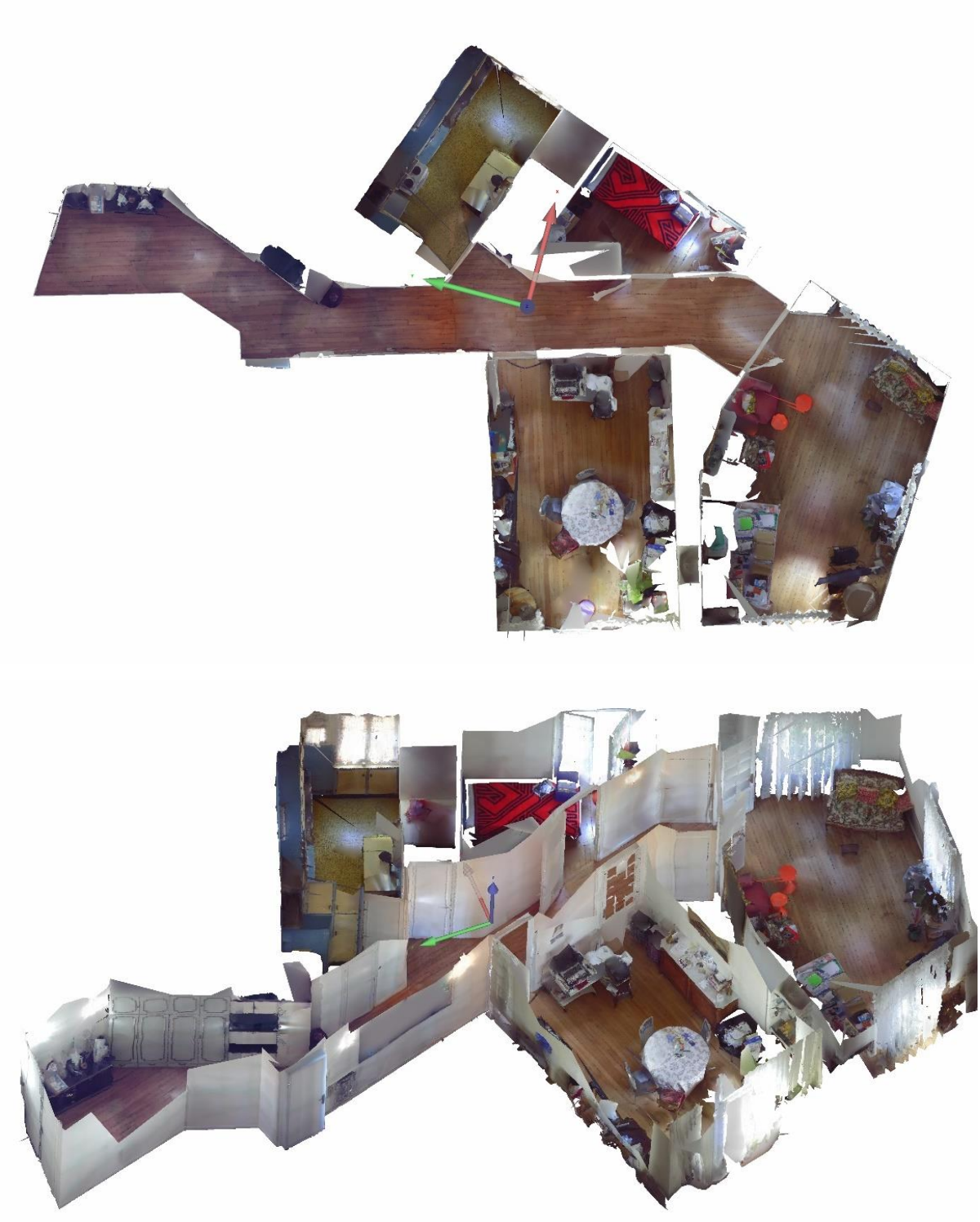


Figure 34. Final textured 3D mesh model of PUC3

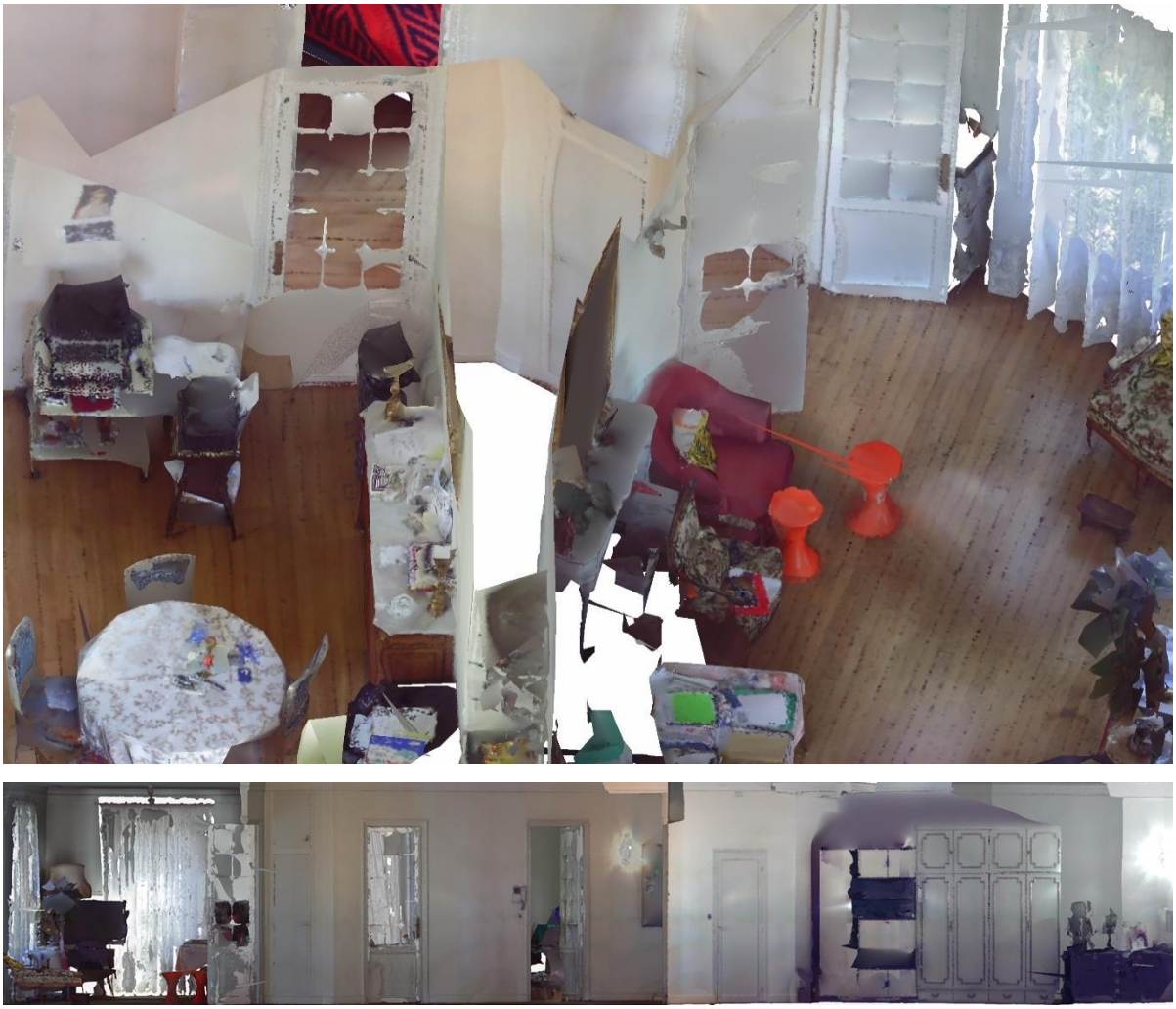


Figure 35. Details of the final textured 3D mesh model of PUC3

A 3D model of a specific piece of furniture was also created from the data collection with a first implementation of the mobile mapping platform. Multiple depth maps captured with a Microsoft Kinect depth sensor were converted to point clouds (Figure 36). These were then registered and combined into a single 3D mesh model (Figure 38). The triangular mesh was then textured from RGB images taken with a separate digital camera (Figure 37).

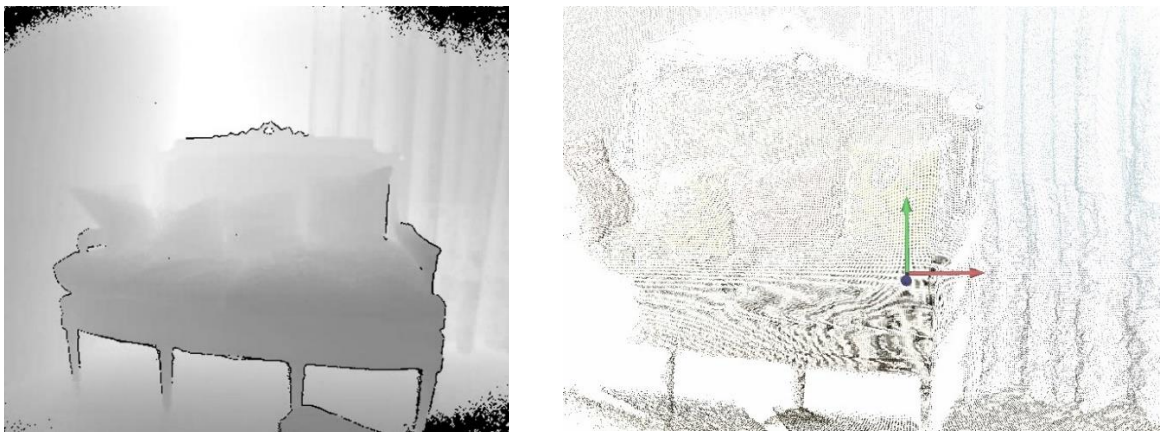


Figure 36. Depth image and respective 3D point cloud from Microsoft Kinect v2 depth sensor

Figure 37. A sample of the images used for texture mapping

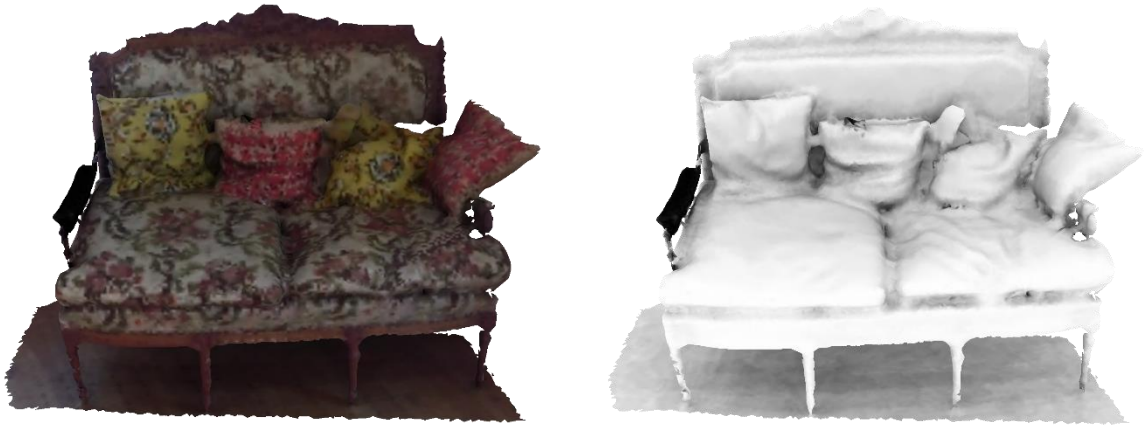


Figure 38. Final 3D model of a historic sofa found in PUC3

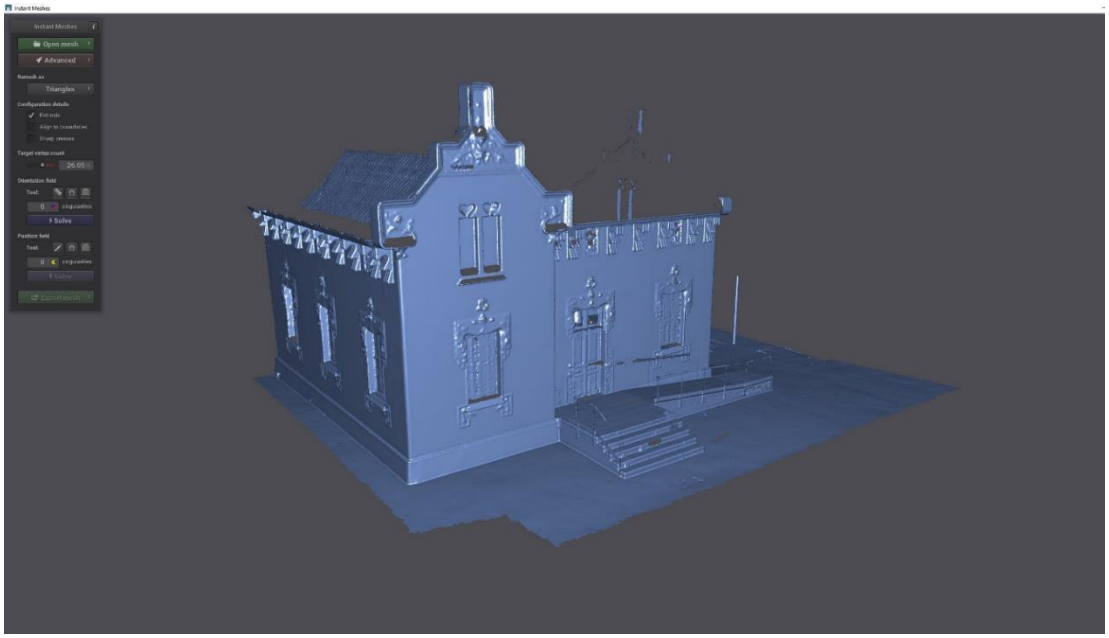
4.4 3D model optimization for VR

The obtained 3D mesh models from almost all of the 3D reconstruction techniques employed in MindSpaces project are often too complicated and complex to be handled directly in real-time Virtual Reality applications. They consist of millions of triangles and multiple texture files are assigned to the geometry to depict all the visual details. To overcome this, a typical workflow from the Computer Graphics Community is adopted. This includes:

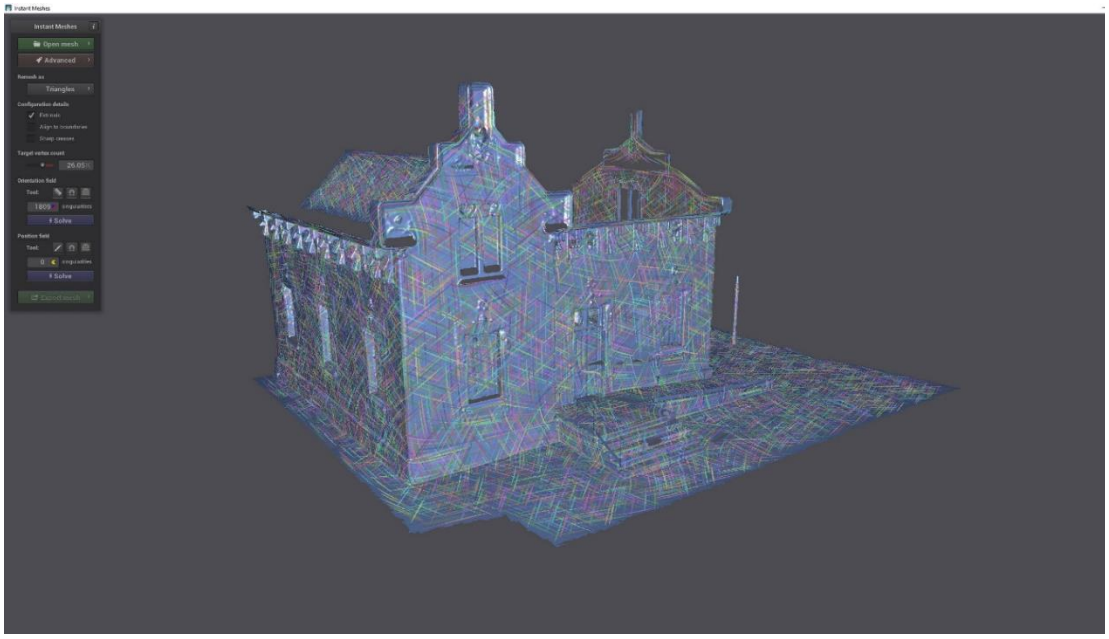
i. Mesh simplification
ii. Physically-Based Rendering (PBR) Texture Maps Baking and
iii. Level-Of-Details (LODs) generation

To simplify the geometry of the reconstructed 3D mesh models “Instant Meshes” tool (Jakob and Tarini 2015) is employed. The original 3D surface is remeshed into an isotropic triangular or quad mesh via a local based approach that optimizes both the edge orientations and vertex positions in the output mesh (Figure 39). The first step computes an orientation field, i.e., a set of directions that the edges of the simplified mesh should align with. The second step computes a local uv parameterization, whose gradient is aligned with the orientation field and which is discontinuous over edges. Finally, a 3D triangular or quad mesh is extracted from the two fields. In this way a simplified version of the original 3D mesh model is created, which consists of less triangles. This is usually referred to as a “low-poly” model.

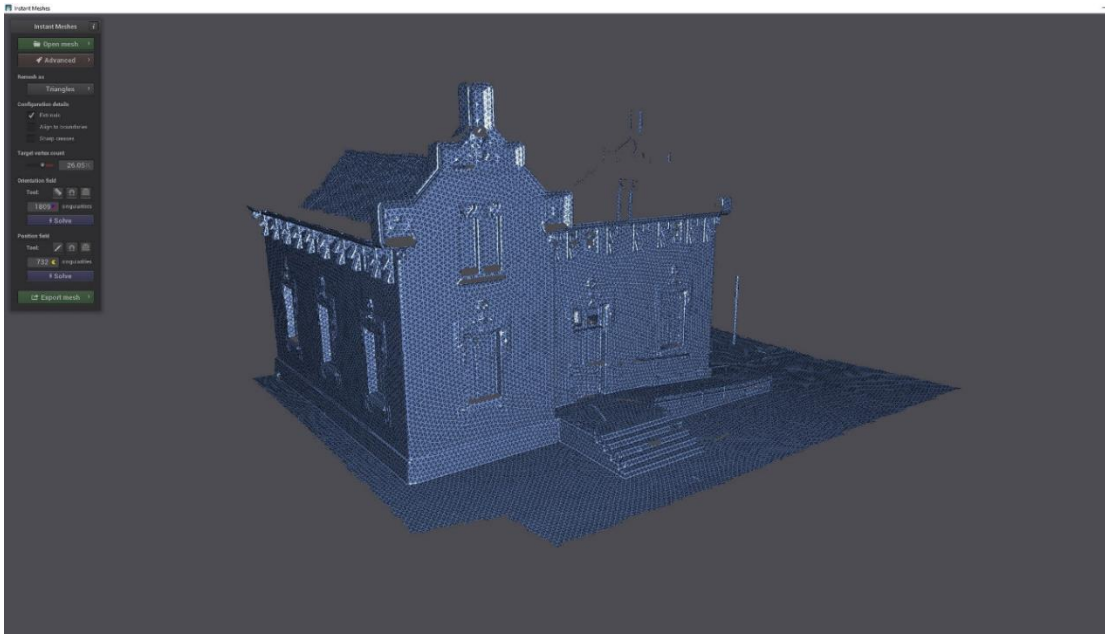
a.


Instant Meshes
Open mesh
Advanced
Triangles
Solve

b.


Instant Meshes
Open mesh
Advanced
Triangles
Solve

c.



d.

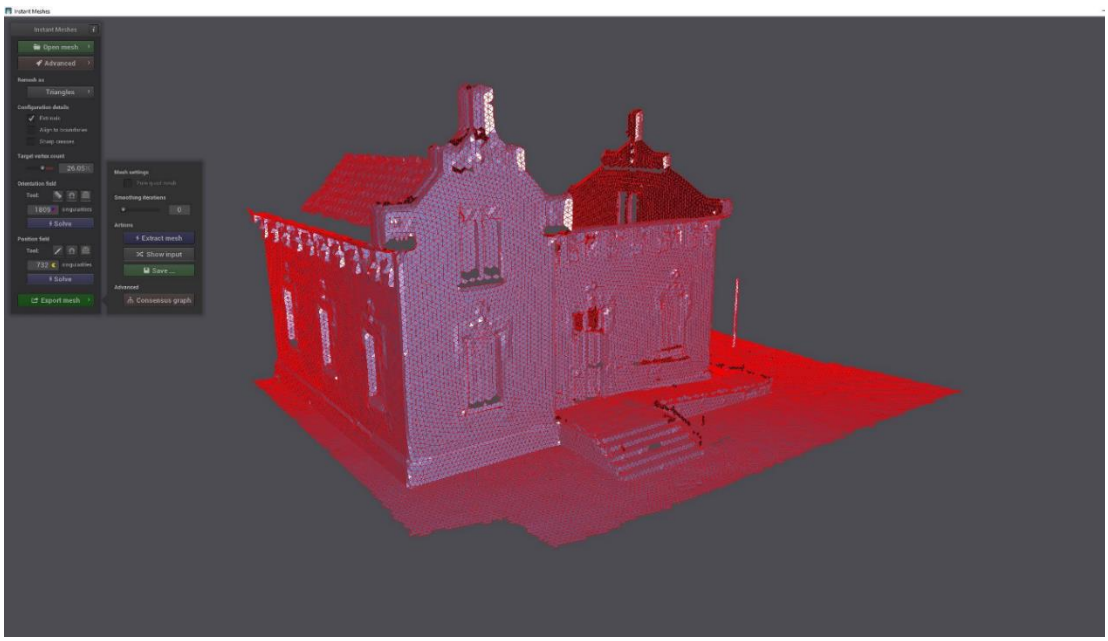


Figure 39. Mesh Simplification by “Instant Meshes” tool. Original Mesh (a). Visualizations of the orientation field (b) and the position field (c). Output simplified mesh (d).

The photo-texture of the original 3D mesh model is then applied to the simplified one by typical uv unwrapping and interpolation techniques. Alternatively, photo-texture can be estimated from the original oriented images. For better visualization in VR, Physically-Based Rendering (PBR) is adopted (Figure 40). Normal Maps, Height Maps and Ambient Occlusion Maps for the simplified 3D mesh are estimated from the more complex geometry of the original mesh using xNormal tool (**https://xnormal.net/**).

a.

b.

c.

d.

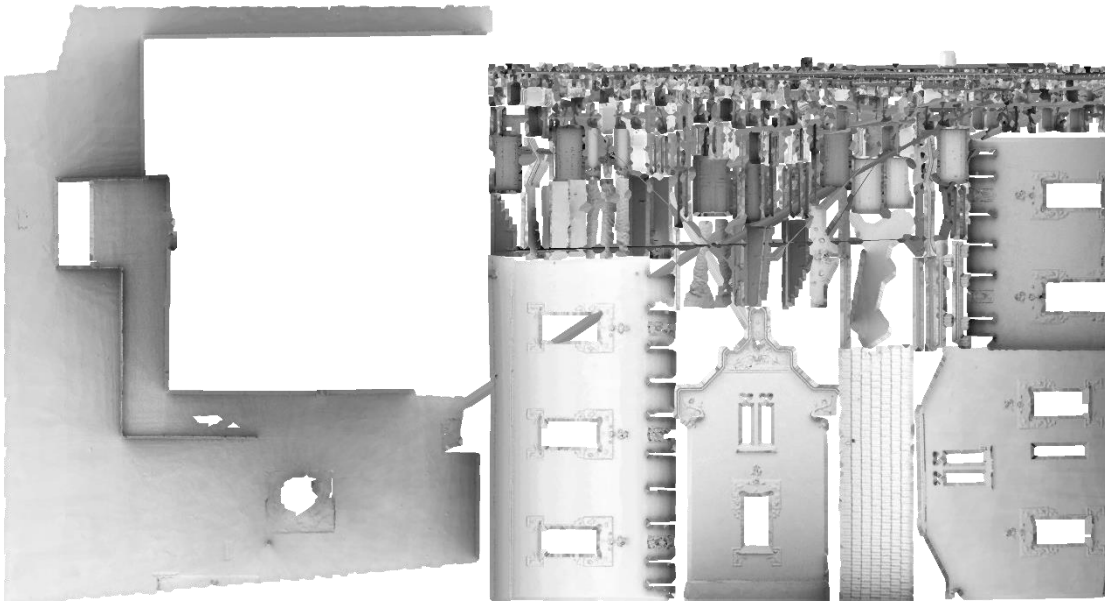
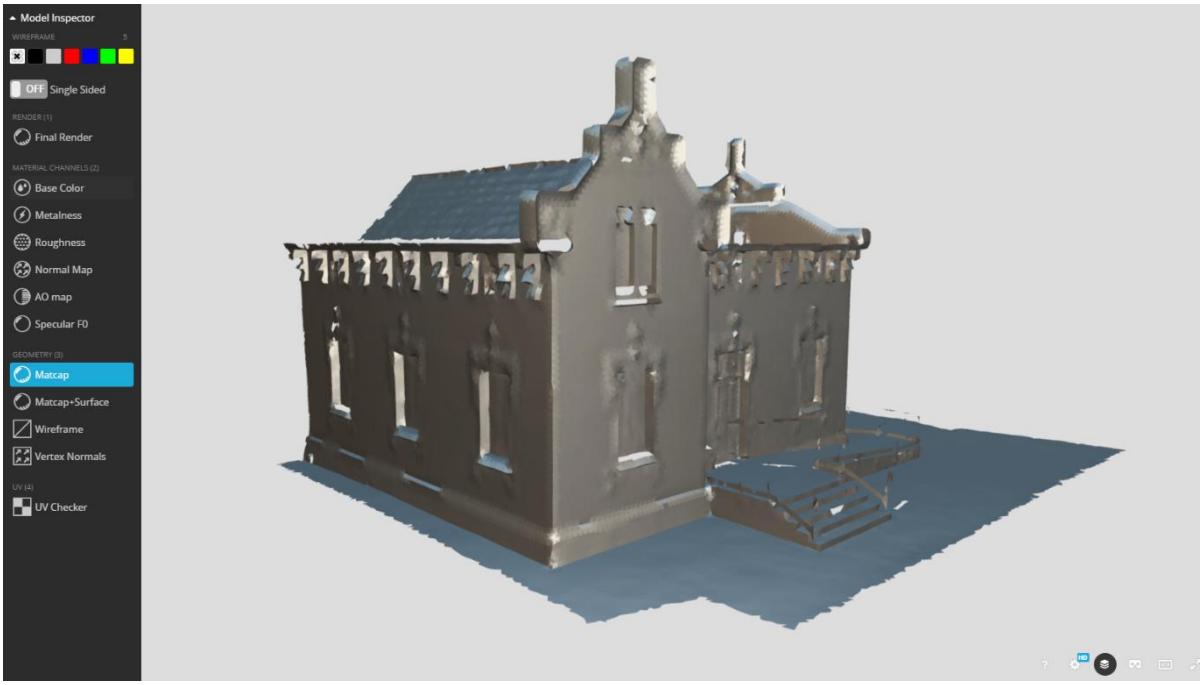


Figure 40. PBR Texture Maps. Texture Atlas (a), Normal Map (b), Height Map (c) and Ambient Occlusion Map (d)


Model Inspector
WIREFRAME
OFF Single Sided
RENDER (1)
Final Render
MATERIAL CHANNELS (2)
Base Color
Metalness
Roughness
Normal Map
AO map
Specular F0
GEOMETRY (3)
Matcap
Matcap+Surface
Wireframe
Vertex Normals
UV (4)
UV Checker

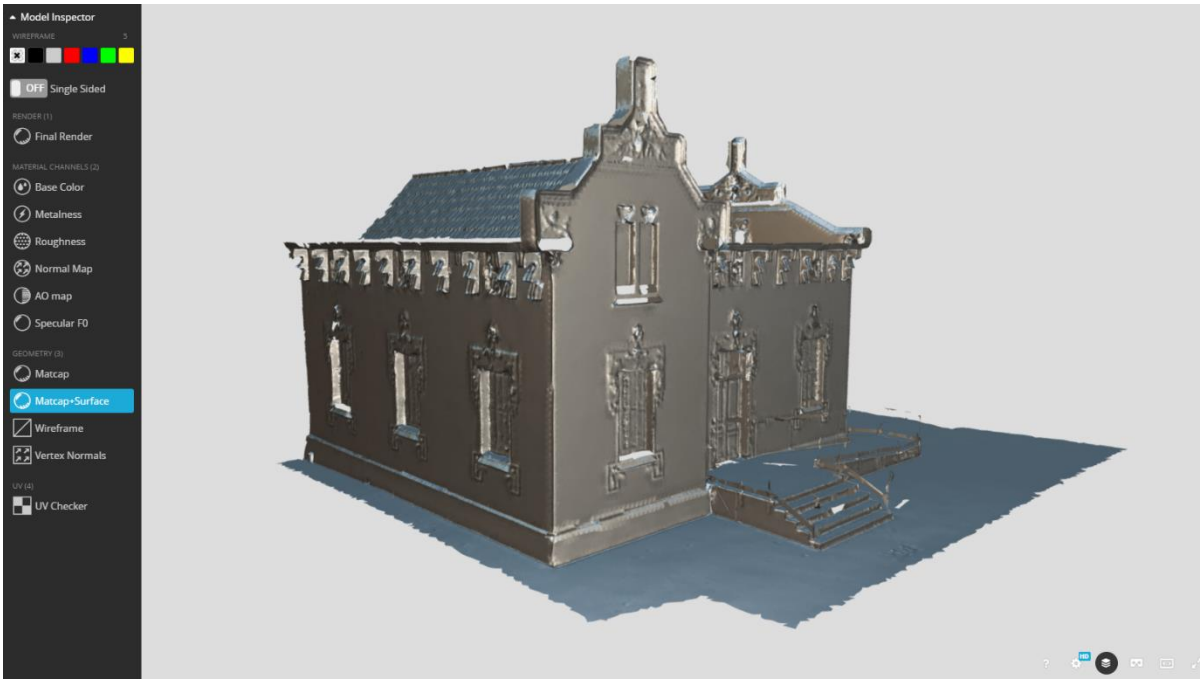
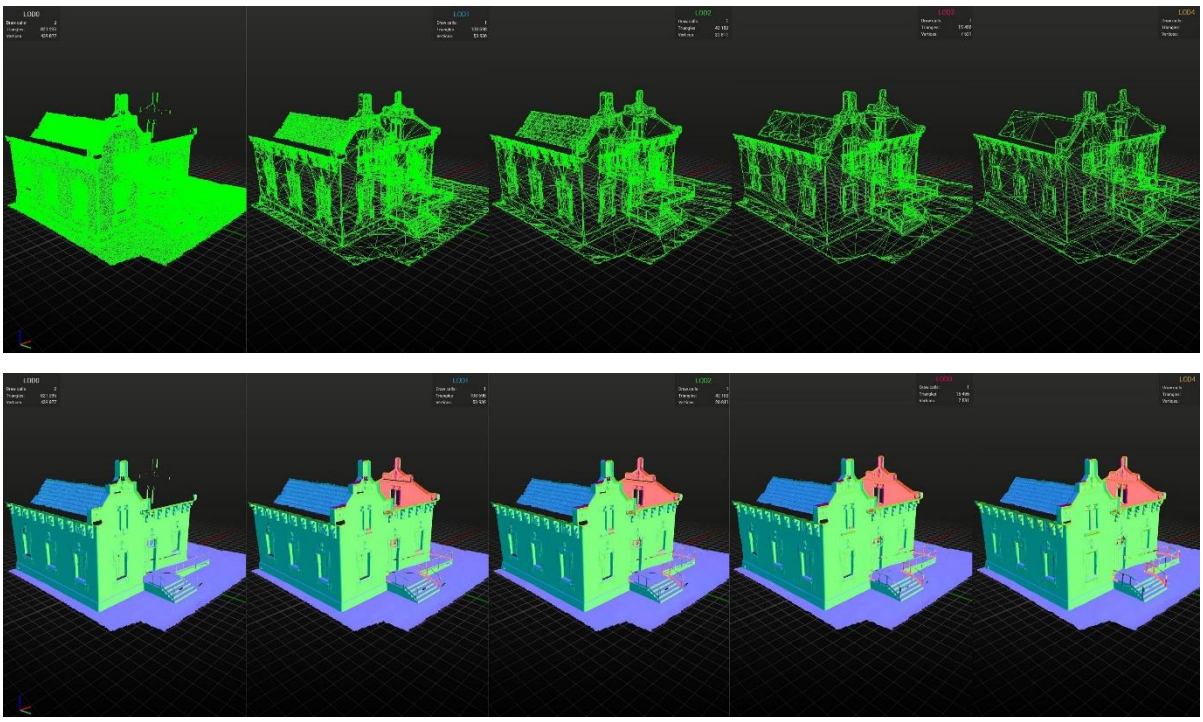

Model Inspector
WIREFRAME
OFF Single Sided
RENDER (1)
Final Render
MATERIAL CHANNELS (2)
Base Color
Metalness
Roughness
Normal Map
AO map
Specular F0
GEOMETRY (3)
Matcap
Matcap+Surface
Wireframe
Vertex Normals
UV (4)
UV Checker

Figure 41. Physically Base Rendering (PBR). Low poly model geometry shown without (upper) and with Physically Based Rendering (PBR). (middle and lower).

Finally, simplygon software (https://simplygon.com/) was used to generate, level of details (LODs) to ease the rendering process in the VR software (Unity 3D) (Figure 42). The latter is essential to achieve real-time performance for exceptionally large meshes.



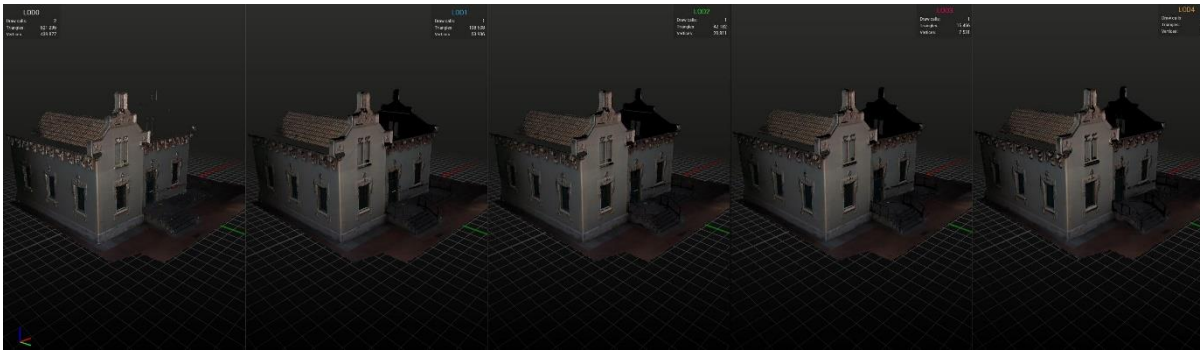


Figure 42. LODs Visualization. Triangles (upper), normals (middle) and final render (lower).

5 AESTHETICS AND STYLE EXTRACTION FROM VISUAL CONTENT

5.1 Colour Palette Generator

This section describes the colour palette generator technique that will be developed in the MindSpaces project along with the Colour Harmony Theory, Models, state-of-the art techniques, and commercial tools. Colour palette generator is a component designed and developed in MindSpaces in the context of WP4 in task **T4.2** - **Aesthetics and style extraction from visual content**.

The scope of our research is to provide a tool and apply techniques for the creation of a colour palette generator aiming to analyse visual content and extract aesthetics for further use from artists and creatives. The aesthetic information and more specifically the colour palettes are extracted from archival materials. The aim is to inspire MindSpaces' users to create novel artworks and designs through the use of the developed colour palette generator.

According to DoA, the colour palette generator is involved in Research Objective RO3 - Design, emotion content extraction and production from multimodal data. The Research Activity RA3.1 Aesthetics and style (design) content extraction from visual data is related to the extraction of aesthetics such as colour themes/palettes from artwork collections that will be integrated into the MindSpaces platform as a basis for inspiring architectural designs and artistic 3D models.

The colour palette generator it is expected to contribute to the following Key Result:

- **KR7. Image aesthetics and style [TRL 5 to 7]:** MindSpaces will apply and extend aesthetics concept extraction on videos and images, based on laboratory validated technologies, so that it can support the needs of a real case scenario.

In the Research Activity RA3.1 our aim is to extract aesthetics from visual content (e.g. abstract paintings) and connect the aesthetic style in an indirect manner, through the colour palettes, to the design process. The methodology that is followed includes unsupervised machine learning technique for the quantization of colour and supervised deep learning architecture for the colour palette generator model.

The following sections include information about colour, colour theories and models, applications of colour palettes and the colour palette generator in the context of MindSpaces project. The last section is devoted to the future plan for the study of colour palettes.

5.1.1 Colour

Everything around us include colour. Colour plays an important role in our daily life and has a strong influence on human behaviour and feelings. According to the study (Khabiri, et al. 2019) the effects of colour is an interesting and challenging topic and there is no "one-size-

fits all" mapping between a colour and the stimuli that can be evoked to a particular person. Colour is studied taking into consideration several aspects and it is recognized that could leave a strong impression on people (Labrecque, Lauren and R. 2012). According to Lennon (Robin and Plunkett-Powell 1997):

"*Colour is the first thing that you perceive when you walk into a room, and it speaks louder than almost any object in a given space*".

Nowadays, technological advances allow to study colour and gain new insights in disciplines such as neuroscience, psychophysics, visual cognition, and biology (Labrecque, Patrick and Milne 2013). Colour-in Context-Theory (Elliot and Maier 2012) is a multidimensional feature. It involves aesthetics and carries meaning. The meanings that colours carry involve two general sources: the societal learning and biology. The context affects the meaning of a colour and therefore colour has different implications for feelings, thoughts, and actions in different contexts.

In conclusion, while colour plays a vital role in our daily life, the effect of individual colours cannot be studied in isolation. Part of people's response to colour is likely to depend on preferences, experiences, and their culture (Labrecque, Patrick and Milne 2013).

5.1.2 Colour Palettes in literature - Related Work

Colour application is a fundamental topic in art, design, and visual media (Haller 2017), (Kita and Miyata 2016). Colour palettes or themes are widely used for sharing colour combinations across multiple applications such as grey-scale photo colorization, image recolouring, colour harmony, and colour palette interfaces and recommendations.

Given the importance of colour, the creation and application of an appropriate colour palette is a key step and challenging task in the design process. The selection of colours for the generation of a colour palette is essential and also involves the colour harmony. Harmonic colours are sets of colours that holds some special internal relation that causes pleasant visual perception. Before the presentation of related work about colour palettes it is appropriate to provide a brief description of the Colour Harmony Theory and Models.

5.1.3 Colour Harmony Theory

In 1672 Newton with his experiments reported the light spectrum. Afterwards several theories were presented and people started systematically explore colour and colour theory. One of the worthy to mention theories is the study from Goethe in the early 19th century. Goethe's approach was more close to art and philosophy than pure science. According to Goethe's words " *'Light and darkness, brightness and obscurity, or if a more general expression is preferred, light and its absence, one necessary to the production of colour . . . colour itself is a degree of darkness.*" colour depends both on light and darkness.

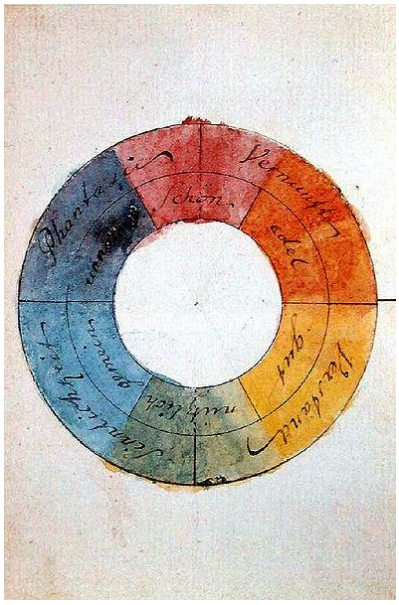


Figure 43. Goethe's famous colour wheel

He created his version of colour wheel and he explored the impact of colour on emotions and assigned different labels to certain colours (Figure 43). In 1905, Professor Albert H Munsell (Birren and Cleland 1969) proposed an orderly system for accurately identifying colours. Rather than describing colours with names which it was considered “misleading” his intention was to create decimal notation. Moon and Spencer Moon and Spencer (Moon and Spencer 1944) in 1944 proposed identity, similarity, and contrast as three principles for colour harmony. According to their study “harmonious combinations are obtained when: (i) the interval between any two colours is unambiguous, and (ii) colours are so chosen that the points representing them in a (metric colour) space are related in a simple geometric manner”. As shown in Figure 44, in Moon and Spencer’s model harmonic colours give sensations of identity, similarity or contrast. Based on their first principle two harmonic colours should not be so close together that there is doubt as to whether they were intended to be identical or only similar. Moreover, based on the second principle pleasure is expressed by the recognition of order. Points in the colour space define a simple geometric space either straight line, circle, triangle, or rectangle.

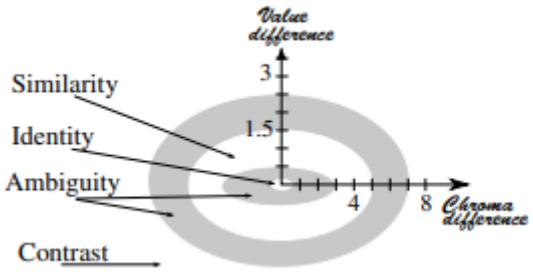


Figure 44. Difference between two colours in the chroma–lightness plane (for a given hue). Harmonic regions are interleaved by ‘ambiguous’ regions

Itten (Itten 1974) proposed a colour wheel (Figure 45). According to Itten in order to examine the expressive properties of a colour you have to relate it to other colours. Itten

formalized theories on the emotional effects of two-color combinations and their properties to generate harmonious artworks. Itten's colour wheel is widely used in art studies and design and it is 12-color circle. Itten introduced the three-color harmony of hues that form an equilateral triangle, the four-color harmony of hues forming a square and the six-colour harmony of a hexagon. Itten was one of the first teachers at Bauhaus school of design. According to Itten "the deepest and truest secrets of colour effect are, I know, invisible even in the eye, and are beheld by the heart alone". Itten's most valuable contribution on present-day colour theory was the association of certain colours with specific emotions. For Itten "Colours are forces, radiant energies that affect us positively or negatively, whether we are aware of it or not". One famous experiment is the "seasons" carried out by Itten. He asked to his students to describe the seasons by selecting a set of colours. To his surprise, all the students used completely different combinations of colours but every student identified which seasons their peers were expressing. His words for this experiment were the following "I have never yet found anyone who failed to identify each or any season correctly… This convinces me that above individual taste, there is a high judgment in man…one which…overrules mere sentimental prejudice".

Figure 45. Itten's colour wheel

5.1.4 Colour Harmony Model -Applications

In literature there are studies for colour harmony models based on the mathematical analysis of user study results. However, these models cannot be considered as generalised models since the size of participants is limited to fewer than one hundred and the evaluation of colours is examined only for two or three colour combinations (SZABÓ, BODROGI, and SCHANDA 2010), (OU, CHONG, et al. 2011), (OU, RONNIER, et al. 2012), (OU and LUO, A colour harmony model for twocolour combinations 2006). (OU και LUO, A colour harmony model for twocolour combinations 2006) One notable study which examines colour themes/palettes and includes 327,381 human ratings of 22,376 colour themes (palettes) is the study from O'Donovan et al (O'Donovan, Agarwala and Hertzmann 2011). Based on this dataset a machine learning model was trained in order to rate colour palettes.

Colour harmonization is an open problem for designers and scientists. It is well-known that there is no there is no formulation for the definition of a harmonic set. However as we described in the previous section, there are some proposed forms, schemes and relations in

colour space that describe a harmony of colours. There are some studies in literature with different applications which aim to help creatives providing tools such as colour palettes generators in order to enhance their artistic work. In (Cohen-Or 2006) a colour harmonization technique is proposed. An example of their method can be seen in the following Figure 46. Given a colour image, their method finds the best harmonic scheme for the image colours.

Figure 46. The original image and the produced harmonized image based on the foreground colours. The background colours are changed based on the foreground content.

In 2016 an aesthetic rating and colour suggestion for colour palettes was proposed (Kita and Miyata 2016). The proposed trained colour palette rating model takes into consideration human aesthetic preferences while a compatible colour suggestion method, extends a given palette while retaining colour harmony. In the Figure 47 an example of the extension of a given colour palette and the application of new colours to new objects of space is shown.

Another application of colour palettes involves the creation of colour palettes generators as a design tool with GUI. The authors of the paper (Stahlke and Loutfouz 2018) propose Chromotype. Chromotype is a computer assisted design tool for Palette generation. Chromotype, is a generative design tool for the creation of colour schemes based on existing palettes, user defined colour sets, and reference images.

A different application of colour palettes involves the extraction of colour palettes from visual data. In the paper proposed by (Afifi 2019) the author's approach for colour palette extraction is based on two-stage clustering. At the first stage, the use of k-means-based clustering is used in order to reduce the number of data points in the given image. Then, a second stage follows using a density-based clustering.

Figure 47. For a bedroom with an associated colour theme (left), a window blind and a sofa with colours assigned according to the extended palette are added to the room.

5.1.5 MindSpaces Colour Palette generator

In the context of MindSpaces the current version of the colour palette generator involves the selection of the data, the training of the model for the generation of colour palettes and the deployment of the service. The selection of the data for modelling the MindSpaces colour palette generator takes into consideration a list of paintings. More specifically, in our approach the goal is to extract the aesthetic style of a set of colour palettes extracted from paintings and then train a deep learning model in order to transfer the knowledge and create a colour palette generator which produces colour palettes with similar aesthetic style.

For the paintings a subset of WikiArts emotions dataset (Saif and Kiritchenko 2018) is utilized. The WikiArts emotion dataset includes a set of 4,105 pieces of art (mostly paintings) that has annotations for emotions evoked in the observer. The pieces of art are from WikiArt.org's collection for twenty-two categories (impressionism, realism, etc.) from four western styles (Renaissance Art, Post-Renaissance Art, Modern Art, and Contemporary Art). The main reason for the selection of the above dataset is the connection of paintings to the emotion evoked to the observers.

For our study from the initial dataset of 4105 paintings a set of 597 abstract paintings were examined. In our research our goals are two-fold. First we would like to create a colour palette generator as a design tool which helps the designers to use a harmonized palette. Second we are also interested to go a step further and find a way to connect the produced colour palettes to emotions and investigate the application of them to the design of spaces. For this reason only the abstract paintings are selected and the extracted colour palettes are connected in an indirect manner to the emotions evoked from paintings. For the extraction of colour palettes the K-means clustering algorithm (MacQueen 1967) was applied. For each abstract painting a 5 colour-palette is extracted and related to an emotion. 597 5-colour palettes are generated. Figure 48, Figure 49 and Figure 51 show a subset of the first 20 produced 5-colour palettes from abstract paintings and the related emotion.

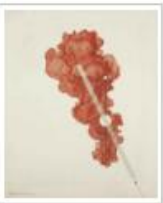
Name	Style	Genre	Artist	Emotion	Emotion Category	Colors	Palettes
Opus 14	abstract art	none	victor_servranckx	surprise	other or mixed		
Fifth Image for J	lyrical abstraction	none	bernard_cohen	other	other or mixed		
It Was Yellow and Pink II	abstract expressionism	none	georgia_o'keeffe	happiness	positive		
Monhegan Island Seascape	abstract expressionism	none	ralph_rosenborg	other	other or mixed		
Invocation I	abstract art	none	theodore_roszak	other	other or mixed		

Figure 48. 5-Colour Palettes produced from abstract paintings from WikiArts Emotions (1-5)

"Meem, Tha, Alif, Lwaath (Heritage) "	lyrical abstraction	none	ali_omar_ermes	anticipation	other or mixed		
Riviera Winter Colour Symphony	abstract expressionism	none	jock_macdonald	surprise	other or mixed		
Untitled	lyrical abstraction	none	sam_francis	anticipation	other or mixed		
Le tumulte à la miâchoire crispée	abstract expressionism	none	marcel_barbeau	fear	negative		
For the Blue Sons of the Air	lyrical abstraction	none	sam_francis	anticipation	other or mixed		

Figure 49. 5-Colour Palettes produced from abstract paintings from WikiArts Emotions (6-10)

Devil Fish	abstract art	none	alexander_calder	surprise	other or mixed		
Matter of Identity I	lyrical abstraction	none	bernard_cohen	surprise	other or mixed		
Komposition	abstract art	none	gosta_adrian-nilsson	surprise	other or mixed		
Claire de Lune	lyrical abstraction	none	roger_bissière	anticipation	other or mixed		
Basel Mural I	abstract expressionism	none	sam_francis	surprise	other or mixed		

Figure 50. 5-Colour Palettes produced from abstract paintings from WikiArts Emotions (11-15)

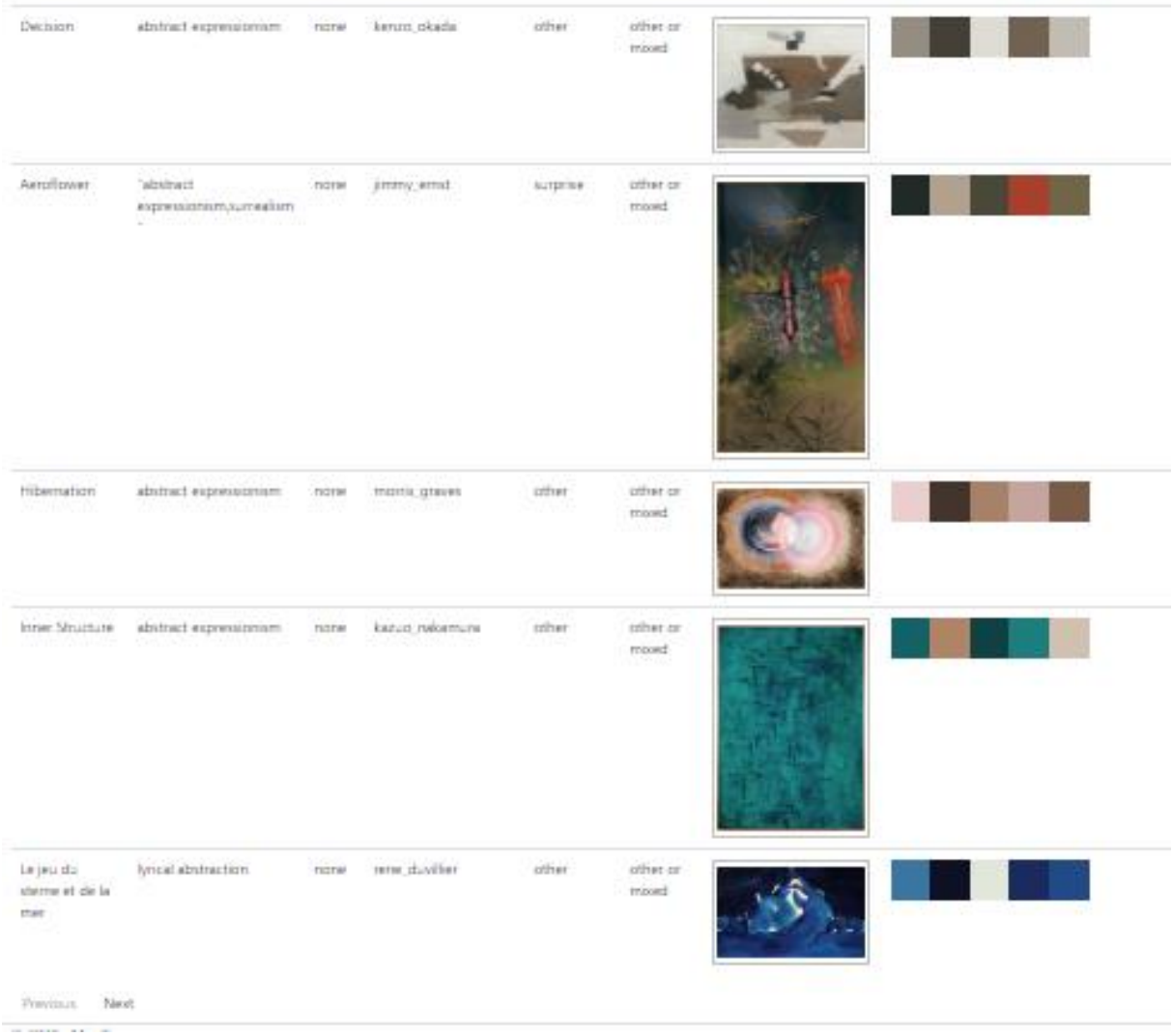


Figure 51. 5-Colour Palettes produced from abstract paintings from WikiArts Emotions (16-20)

The proposed tool is a colour scheme generator which uses deep learning technique in order to learn colour styles from visual data such as paintings, movies or photographs. If the user does not provide a colour the colour palette generator creates a 5-colour palette. The user may also add up to 4 colours in any order and the colour palette generator will fill in the missing colours.

To support the colour palette generator multiple models (30) have been trained. The formula for combinations is:

nCr = n! / r! * (n - r)!, where n represents the number of items, and r represents the number of items being chosen at a time. More specifically:

- for the case of a single locked colour from a user we trained 5!/1!*4!=5 models

- for the case of two locked colours from a user we trained 5!/2!*3!=10 models

- for the case of three locked colours from a user we trained 5!/3!*2!=10 models

- for the case of four locked colours from a user we trained 5!/4!*1!=5 models

The pix2pix deep learning architecture is used. Pix2Pix is a Generative Adversarial Network (GAN) model designed for image-to-image translation applications. This deep learning architecture was introduced by (Isola 2017) in 2016. The GAN architecture comprises a generator model for producing new synthetic images, and a discriminator model that classifies images as real or fake. The discriminator model is updated directly, whereas the generator model is updated via the discriminator model. The two models are trained simultaneously in an adversarial process where the generator seeks to better fool the discriminator and the discriminator seeks to better identify the counterfeit images. The Pix2Pix architecture is a conditional GAN since the generation of the output image is conditional on a source image. The discriminator is provided both with a source image and the target image and must determine whether the target is a plausible transformation of the source image. The generator is trained via adversarial loss, which encourages the generator to generate synthetic images in the target domain. The generator is also updated via L1 loss measured between the generated image and the expected output image. This additional loss encourages the generator model to create synthetic translations of the source image.

5.1.6 Future Work

The first version of the colour palette generator is based on a set of abstract paintings. In the future work we may also try to include data from other sources such as movies or photographs, artworks or images related to indoor design. Moreover, the colour palette generator is fixed and produces a set of 5 colour combination. This could also be extended. We could also create a rating model for the produced colour palettes and may include it during the extraction of colour palettes from the trained data to improve the order of colours. An alternative also connection of colour palettes to emotions could be also considered.

5.2 Style Transfer

5.2.1 Connection to MindSpaces HURs

Style transfer is a WP4 component and specifically of the task T4.2 which refers to the aesthetics and style extraction from visual content tasks. This task addresses the user requirements mentioned in D7.1 associated with the texture and materials' proposal module of MindSpaces, namely the UR_10, UR_70, UR_77. Regarding UR_10, style transfer module will provide to the system a set of produced materials (a library of materials) in order the corresponding module to achieve the correlation with the emotional state, as requested. In UR_70, an architect wants to be able to create novel and inspiring textures based on famous paintings and artwork. Style transfer framework will produce materials based on a pre-set group of famous artwork where the user will be able to choose unique and inspiring textures that do not exist in current 3D modelling market. Concerning UR_77, an architect wants to have the option to define/change the material of an urban/interior space. This will be easily feasible, through the material library that will be created through the style transfer module.

Table 7 Relevant user requirements reported in D7.1 for Style Transfer on materials

User Requirement (UR)	Associated HLUR	Detailed description	Functional or Non Functional (FR/N-FR)	Priority based on MoSCoW framework
UR_10	HLUR_2	As an architect I want to correlate material palettes and colours with the emotional state of users in designed workplace environments	Non Functional	M
UR_70	HLUR_11	As an architect/Designer I want to be able to create novel and inspiring textures that are based on the aesthetics of famous paintings and other images of artwork that do not exist in current 3D modelling market.	Functional	M
UR_77	HLUR_13	As an architect I would like to be able to define/change the material of an urban/interior space in MindSpaces platform	Functional	M

5.2.2 Style Transfer in General

Neural style transfer is a deep learning technique applied to images and videos in order to compose an aesthetically modified output, in the style of a style reference input. The content image/video along with the style input are blend together through a convolutional neural network-based architecture and a stylized output is produced, preserving the main content but changing the style of the initial visual input. This application is extensively used in artistic and design domains, architecture, and game development industry. In Figure 52, a style transfer example is illustrated. A photo of Chicago city (middle) is combined with one of the most famous paintings of Katsushika Hokusai, “The Great Wave off Kanagawa” (left), to produce a new image of Chicago city in the style of Katsushika Hokusai (right).

Figure 52. Transferring the style of the famous painting “The Great Wave off Kanagawa” of Katsushika Hokusai in a photo of Chicago city.

5.2.3 Related Work (State-of-the-Art)

In the seminal work of (Gatys 2016), the authors present for the first time, high-quality style transfer results. Their technique demonstrates the ability of Deep Neural Networks (DNNs) to encode not only content but also the style information of an image, by exploiting the second-order statistics in Gram matrices and capturing the linear dependencies between several feature vectors. The proposed framework is based on an iterative optimisation process where the image is updated in each iteration in order to minimise the loss between content and style, which is calculated by a loss network. In (Huang and Belongie 2017), the authors introduce a new technique which qualifies for real-time arbitrary style transfer. They adapt the channel-wise statistics of content features with the statistics of the style features by using a simple adaptive instance normalization layer (AdaIN). An encoder-decoder architecture is adopted, where the encoder is consisted of the first few layers of a pre-trained VGG-19 network. AdaIN receives a content input and a style input. After encoding them in feature space, both feature maps are fed to an AdaIN layer that aligns the mean and variance of the content to those of the style, producing the target feature maps. It should be mentioned, that AdaIN has no learnable affine parameters. The authors in (Ulyanov, Vedaldi and Lempitsky 2017) demonstrate a new module based on instance normalisation, as a replacement of the batch normalisation, that leads to the improvement of the performance and the quality of image styling. Additionally, they propose a new learning formulation where the training generator network samples uniformly the set, resulting to the high fidelity and diversity of the produced stylized outputs. Their method takes feed forward texture synthesis and image styling closer to the quality of generation via optimisation while retaining the speed advantage. The authors in (Sheng, et al. 2018) introduce a new method, using a style decorator for semantic style feature propagation and an hourglass network for multi-scale holistic style adaptation. Moreover, they integrate the Zero-phase Component Analysis (ZCA) operation in their style transfer method. The VGG network is utilised to extract image features, and ZCA is used to project content and style features into the same space. Then, the transferred features are reassembled by a patches-based operation. In the last step, the features are reconstructed through a decoder network and the output styled

image is produced. The authors in (Xu, et al. 2018) trained adversarially a feed-forward network in order to achieve arbitrary style transfer. They proposed techniques to tackle the problem of adversarial training from multi-domain data. In adversarial training the generator and the discriminator (both consisted of conditional networks) are updated alternatively. The generator is trained to fool the discriminator, as well as to satisfy the content and style representation similarity to inputs. The generator is built upon the previous work (Huang and Belongie 2017) for arbitrary style transfer and the discriminator is conditioned on the coarse domain categories, which are trained to distinguish the generated images from the same style category. Furthermore, they demonstrate a mask module to control, automatically, the level of styling by predicting a mask to blend the styled and the content features. Finally, they exploit their trained discriminator to rank and find the representative generated images in each style category.

The style transfer problem becomes even more challenging when the style input is based on a collection of images such as paintings of a creator (the style of the creator) or the style of a school of art (impressionism, cubism etc.). To tackle this problem the authors in (Zhu, et al. 2018) introduce a new model that extends the architecture of Generative Adversarial Networks (GAN), known as Cycle-Consistent Adversarial Network or CycleGAN. This model allows the translation of an image from a source domain to a target domain without the need of paired examples. In order to achieve this, their approach utilizes two mapping functions and the corresponding adversarial discriminators. Furthermore, they learn an inverse generator which produces an image identical to the original content one using as input the styled image. In (Chen, 2018) the authors propose a novel adversarial network, knows as Gated-GANs, that is able to transfer multiple styles in a single model. The main components of their network are: an encoder, a gated transformer, and a decoder. The function of the gated transformer is to allow the user to select the style that will be applied, by switching gate. Gated-GANs are trained for multiple styles in order to generate new styled images through weighted connections between the branches of the gated transformer. The authors in (Sanakoyeu, et al. 2018) adopt an encoder-decoder network architecture, proposing their own proposed style-aware content loss. They introduce a fix point loss that ensures styling has converged and reached a fix-point after one feed-forward pass. This style-aware content loss forces the styling to take place in the decoder. In (Liu, Michelini and Zhu 2018) the authors propose the Artsy-GAN, a variation of CycleGAN, that uses the so-called perception loss (Johnson, Alahi and Fei-Fei 2016) instead of the Cycle-Consistency loss. The primary scope of this approach is to introduce a new objective function for diversity in image-to-image translation. The generator of the proposed framework includes three main branches, where each one is fed with the same input and extracts three different channels of the output images: two colour channels and one luminance channel. At the last step, the output image is reconstructed in its final RGB format.

5.2.4 Style Transfer for the production of new materials

The initial version of the style transfer module of MindSpaces is based on AdaIN (Huang, 2017) which is able to transfer arbitrary new styles, in contrast to other models that allow only one style or limited number of styles. The network receives as inputs a content image and an arbitrary style image. An encoder – decoder architecture is utilized, where the

encoder module is adopted from the pretrained VGG – 19 network of (Simonyan and Zisserman n.d.), consisted of its first few layers (up to relu4_1). The initial inputs are mapped through the encoder to the feature space and provided to the AdaIN layer as inputs. The mean and variance of the encoded content are matched with the corresponding style ones and the output is produced.

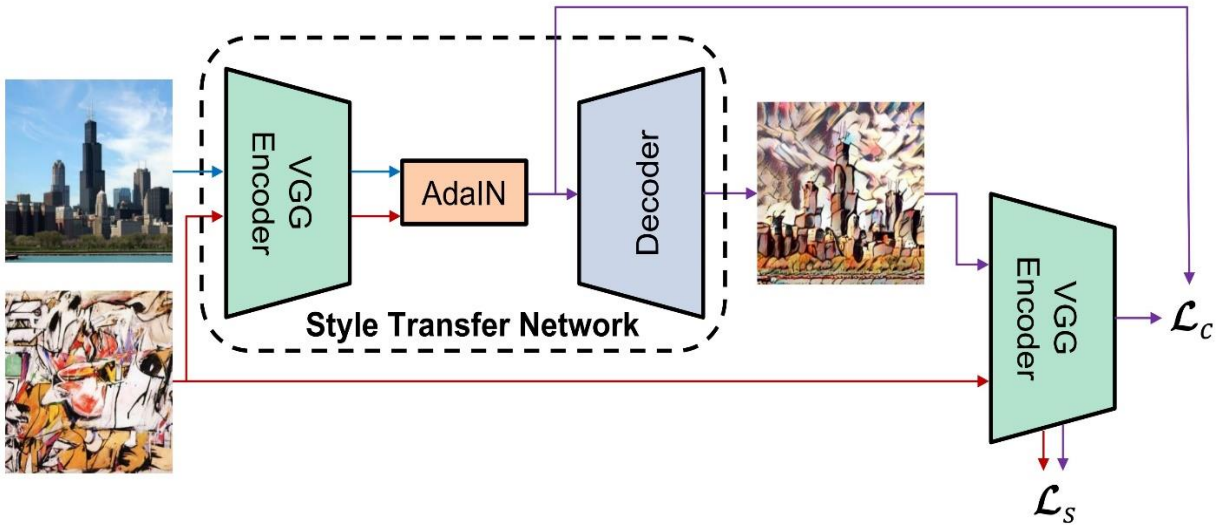


Figure 53. The Style Transfer Framework

On the next step, the previously produced output is fed to a decoder module which decodes it to the image space and the stylized image is created. The decoder mainly reflects the encoder's components, with the pooling layers removed and up-sampling ones used instead. Moreover, it should be mentioned that the decoder module does not include any type of normalization layers in order to achieve image stylization from a wide set of styles.

The loss function of this implementation, used during the training of the decoder, is close to the one described in (Ulyanov, Vedaldi and Lempitsky 2017). Thus, a pretrained VGG – 19 was utilized and the weighted combination of content and style loss was calculated. The total framework of the implemented approach is depicted in Figure 53.

Initially, the network was trained using the Microsoft COCO (MS-COCO) dataset as content images and the WikiArt dataset as style images. MS-COCO includes approximately 80,000 training examples, with the content originating from several fields. The WikiArt paintings dataset is an image collection of 81,472 paintings images, from more than 1,000 artists. This dataset contains 27 different styles and 45 different genres. Based on our knowledge, it is currently the largest digital art dataset publicly available for research purposes.

For further improvement and in order to steer the optimization of the implemented network in the direction of MindSpaces' scope and specifically in materials style transfer, we proceeded to a second round of training. We used as content input a total number of 75 material images provided by the user partners and 53 painting images from several artists as style input. The 53 different paintings used during this training cycle can be observed in Figure 54.

A total of 265 material content images, suggested from the user partners and covering the project's use cases, were included in the experiment with the new trained model. The final outcome was an image library consisted of 14,045 novel and inspiring materials, based on the aesthetics of famous artwork.

For the implementation of the framework and the execution of our experiments, TensorFlow[3] backend and keras[4] deep learning neural network library in Python were used. The experiments run on the NVIDIA GPU GeForce GTX 1080 Ti.

The Style Transfer module has been evaluated with respect to processing time and mainly through the qualitative results which are illustrated in the following section.

In the following table (Table 8) the runtime of the implemented method for Mindspace purposes compared to the baseline is illustrated. Experimenting on different size of images, we see that inference speed of the implemented approach is significantly higher than the baseline's method.

Table 8: Speed in seconds for materials' style transfer implemented approach vs Gatys.

Image size	(Gatys 2016)	Implemented approach
$256 * 256$	$14.19s$	$0.027s$
$512 * 512$	$46.79s$	$0.098s$

[3] https://www.tensorflow.org/

[4] https://keras.io/

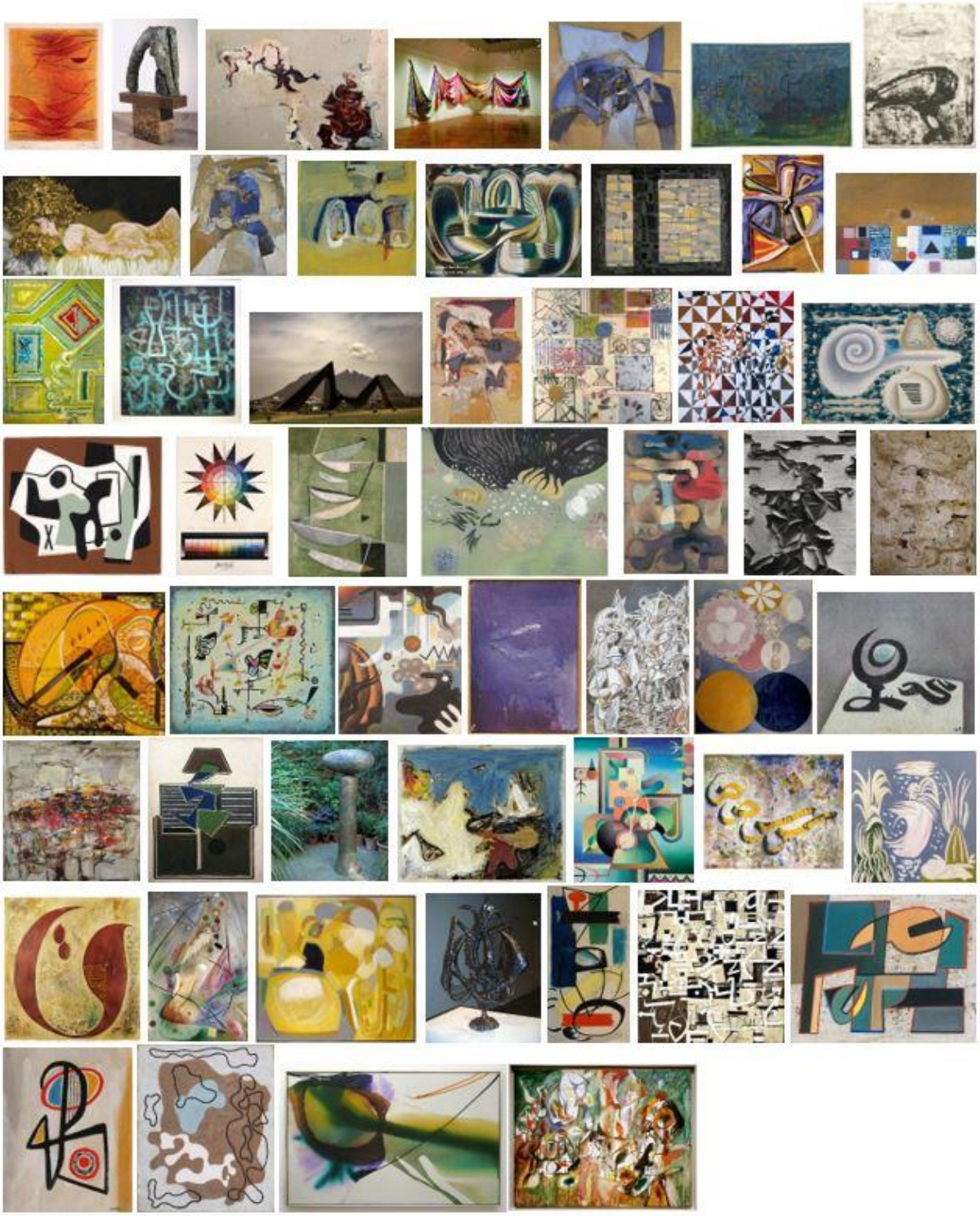


Figure 54. All the 53 paintings used for the training of our implementation

In Figure 55, we demonstrate some qualitative results of the baseline approach along with the corresponding results of the implemented method. The input style image is presented on the top left corner of each example and the content image on the top right. The stylized output of the implemented approach is depicted in the bottom left corner, and the corresponding output of Gatys is shown in the bottom right field. We can observe that the results of the baseline and the implemented method are comparable in terms of quality.

However, we see that on “The Muse” example, the Gatys output seems slightly deformed and some details are missing. On the other example, in the case of the baseline approach we notice that in some sections there is a colour distortion while in the implemented method the colours seem more natural.

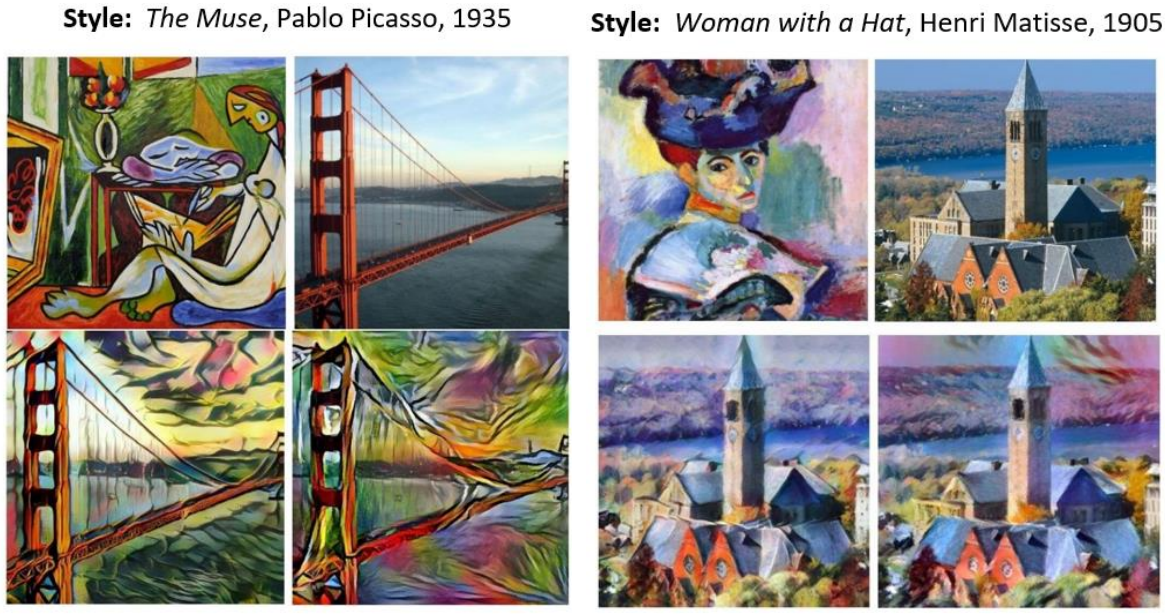
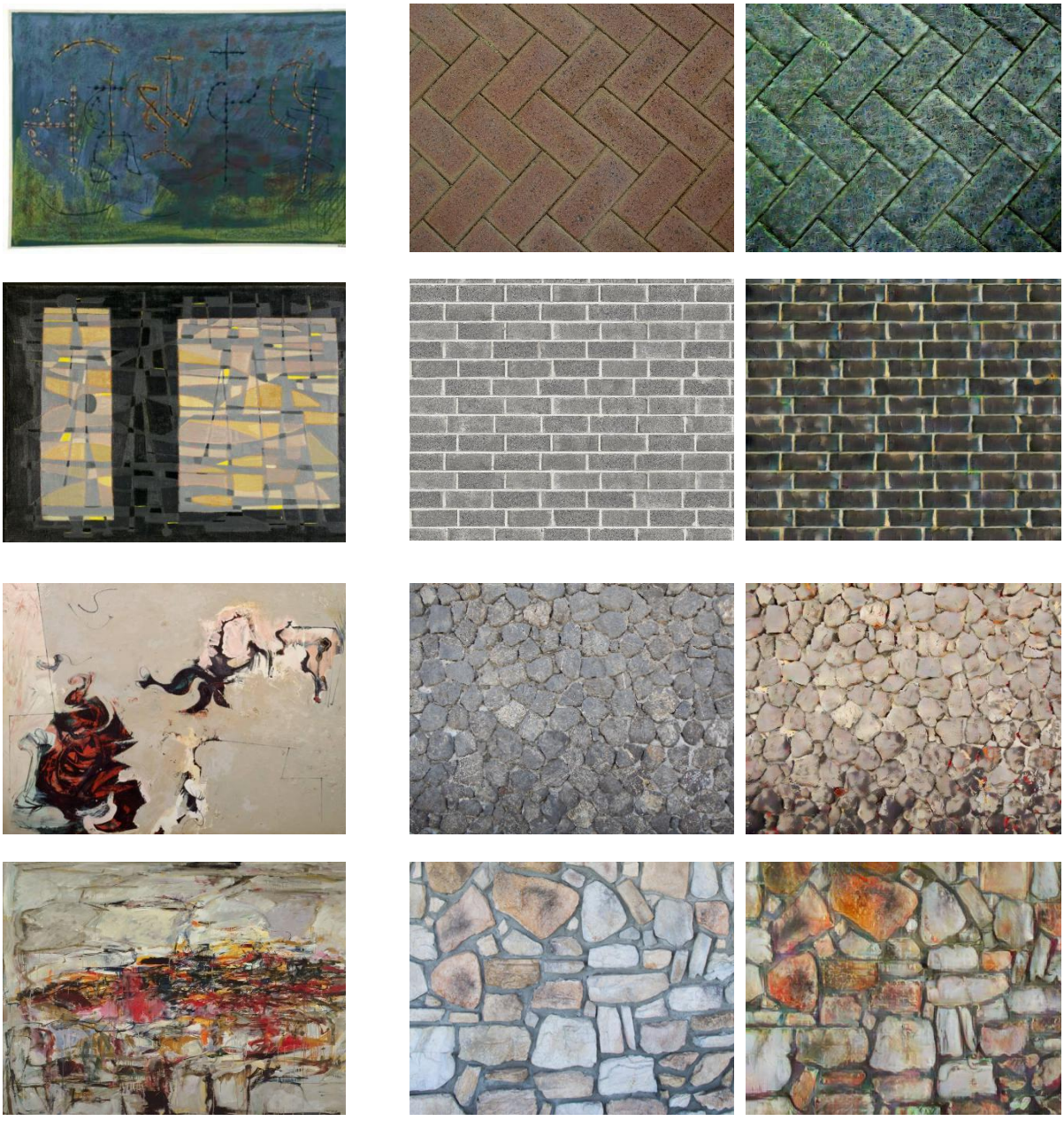


Figure 56. Qualitative comparison of our Style Transfer implementation vs Gatys et al. using as style the painting “The Muse” (left) and the painting “Woman with a Hat” (right).

In this final section, we demonstrate how style transfer can transfer the paintings’ style to materials, so that it can be become clearer how target images can change and provide the appropriate input to 3D reconstruction and re-texturize the 3D models appropriately. The primary objective of our algorithm here was to transfer the features from the paintings to the appropriate regions in the target material images. In Figure 57, we demonstrate a small subset of our materials library. The style image (artwork) is presented on the left column, whereas the content image and the stylized output material are in the middle and the right column accordingly.



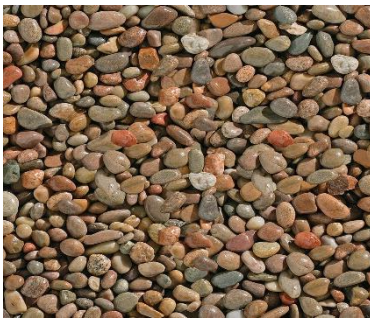
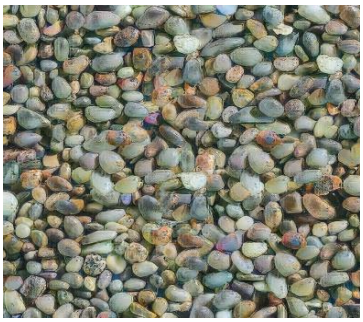
Style Content Output





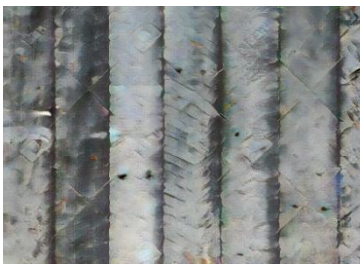




Style Content Output





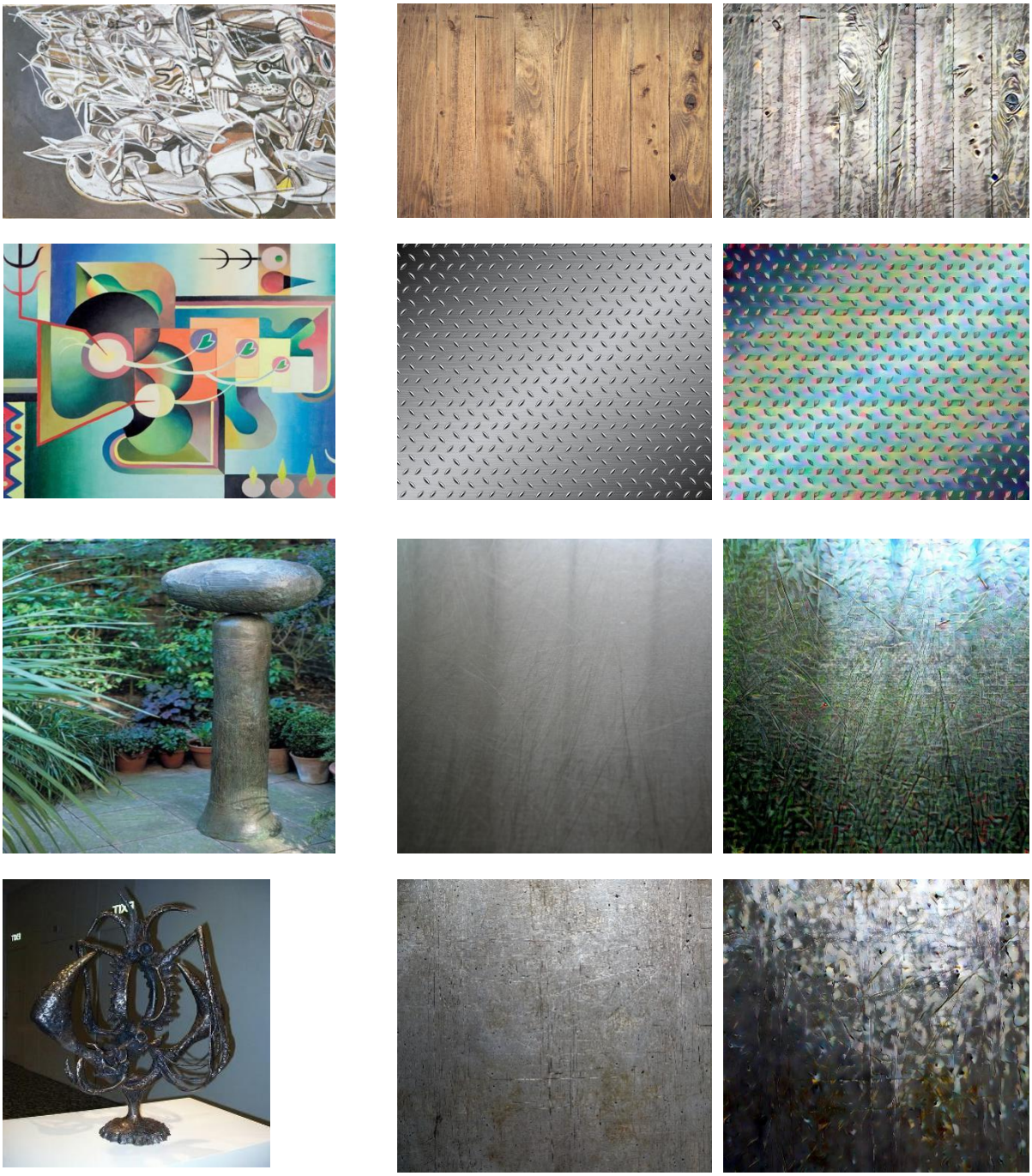


Figure 57. Style Transfer from famous paintings to wooden and metal materials

5.2.5 Pipeline of the aesthetics services

Table 9. Main functions of the aesthetics services

Function Name	Description	Data input *(expected from other services)*	Data output *(to other services)*
Style Transfer	This function transfers the style of one of the 53 available styles of paintings to a content image for the creation of a new version of the	The content image and the ID of the style image	The stylized image (new material) saved in the File storage

	initial material stored in the File storage		
Colour Palette (API)	Returns in JSON format the colour palettes extracted from paintings along with the related attributes.	A call (query) to the API. http://160.40.52.102:14140/api/colorpalettes	A JSON format output which includes the list of colour palettes
Colour Palette Web UI	A GUI for the examination of colours from colour palettes	A request to the following address http://160.40.52.102:14140/list	A GUI with all the details of the created colour palettes

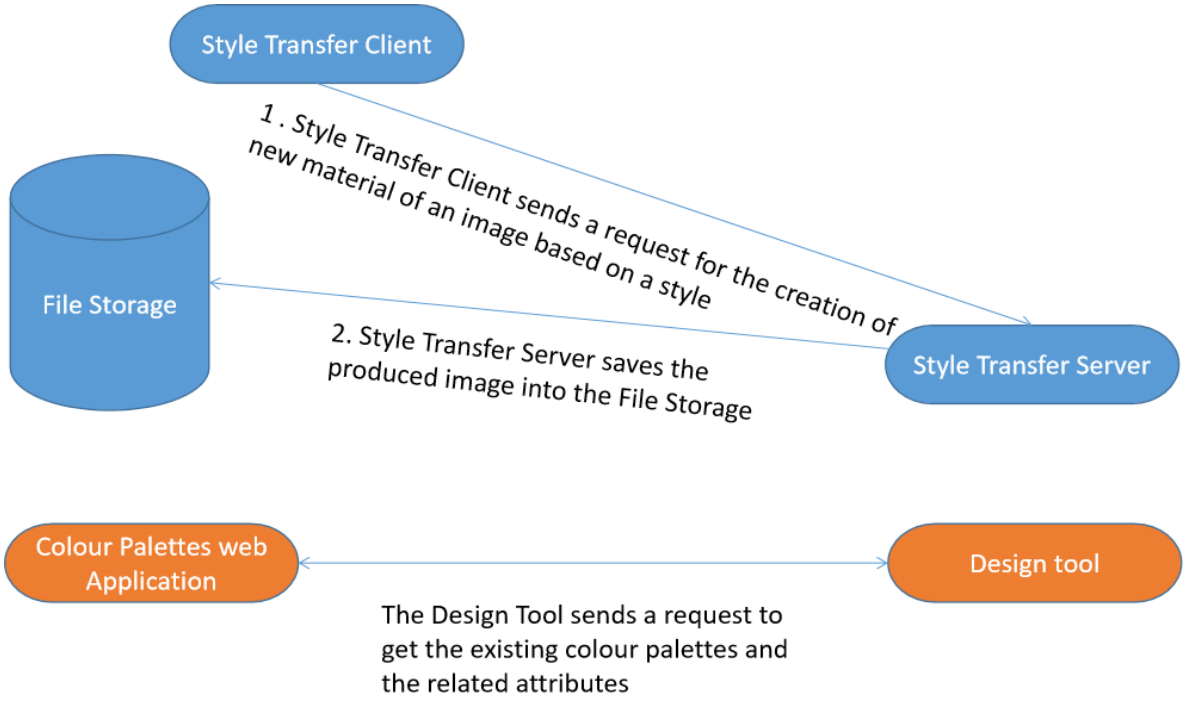


Figure 58. Logical design of aesthetic services

5.2.6 Future work on new materials' production through style transfer

As next steps, we plan to propose a novel framework focused on style transfer for fast and efficient unique materials' production. There are different possible options regarding the implementation of such a framework in terms of the style type that will be transferred. Some of them include the combination of a content image and more than one styles, others the combination of a content image and a collection of several artwork created from a specific artist and some other focus on the combination of a content image with another similar non art-related image in order to obtain more real-life produced output (i.e. utilizing as content a brick material image and as style another brick content image which results to a great unique very realistic output). In our future work, we may take into consideration and experiment with all these approaches in order to examine which one is able to generate new materials with the best visual content and highest level of realism.

6 TEXTUAL ANALYSIS

6.1 Task definition

Text is one of the sources of signals relevant to the purposes of the project as it contains insightful information to be used as a basis for reasoning for art and design production. The task covers the development of tools for processing the textual data and capturing its underlying semantics for a number of applications: (i) analysis of social media texts to identify emotions, opinions, and needs of citizens on some particular topics related to outdoors design; (ii) analysis of special discourse resources related to office spaces for detection of sentiment and opinions on indoors design-related aspects; and (iii) analysis of transcribed dialogues with seniors on living space design to identify their feelings and opinions on some particular desired functionality that would satisfy their needs.

The generic textual analysis in MindSpaces is addressed as a sequence of steps: (i) morphological analysis, (ii) syntactic analysis, (iii) semantic analysis. The output of each analysis is feeding the next step, and the level of abstractness required in the semantic structures can be attained gradually. Sentiment analysis is performed in addition in parallel to the generic pipeline. The results of both subcomponents are merged into a single formal text representation at the end.

The position of the textual analysis module in the MindSpaces Platform is shown in Figure 59. There are two input points to the module: (i) through a Listener that connects to Crawling and Scraping services and receives textual data on a regular basis, (ii) through a special user interface (Web GUI) where designers can submit individual texts. An output distributor propagates results further to other MindSpaces modules depending on the input. It implements the gRPC[5] client and sends the result structures to Storage and Knowledge Base (first, passing them through Formal Text Representation module) via gRPC communication or returns results directly back to the Web GUI.

[5] https://grpc.io/

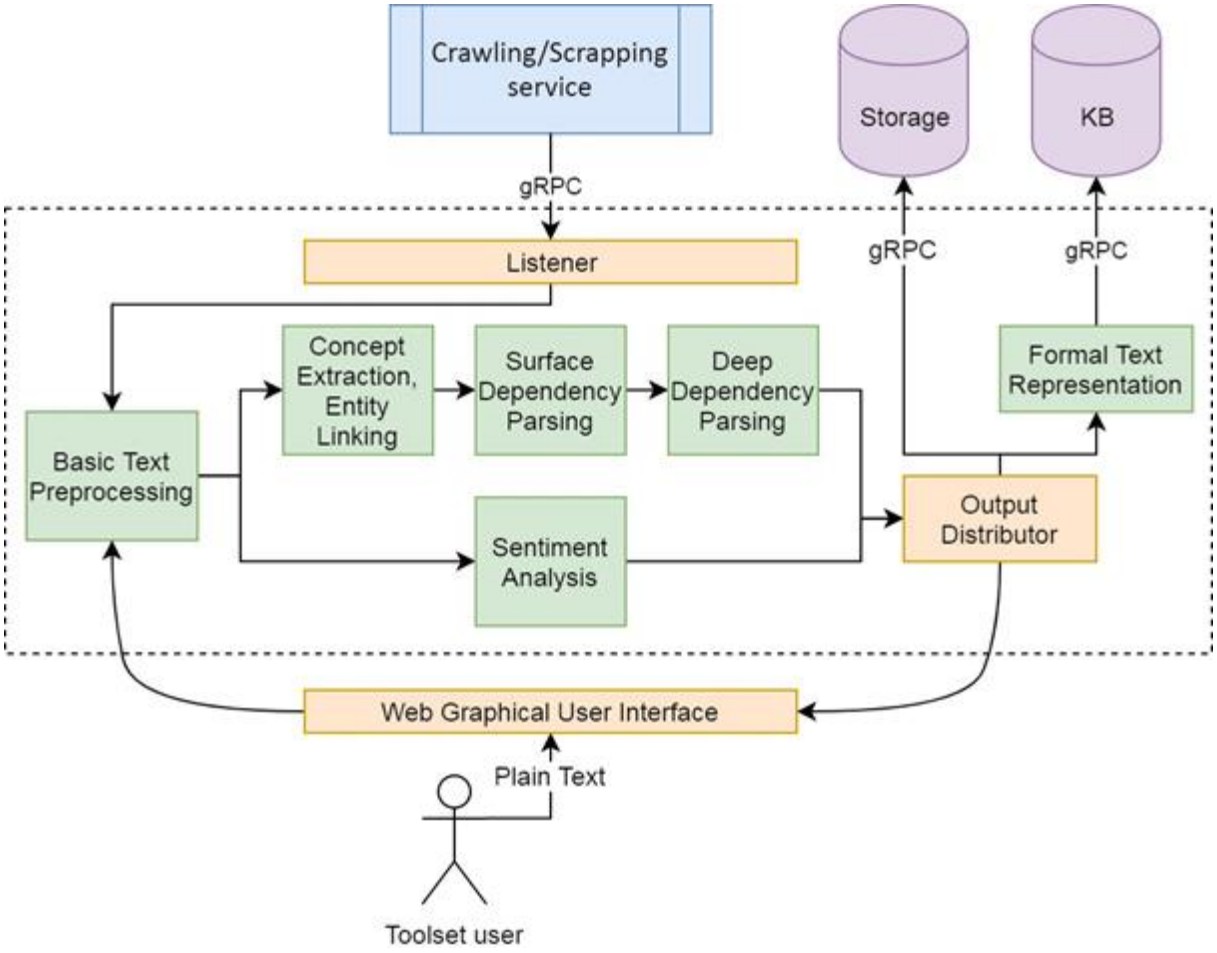


Figure 59: The scheme of textual analysis module within the MindSpaces Platform

During the first half of the project, the work on Textual Analysis has been focused on:

- making an overview of (i) the annotated corpora;(ii) the available open-source tools that can be useful for the analysis pipeline; (iii) existing tools for concept extraction;
- training statistical modules for the first steps of the analysis;
- developing new graph-transduction grammars to be used for multilingual semantic relation extraction between the words of the sentence for three of the five languages of the project: English, Spanish and Catalan;
- adapting existing models for sentiment analysis for three of the five languages of the project: English, Spanish and Catalan;
- providing a first evaluation of the modules used in the analysis (i.e. concepts extraction, dependency parsing, sentiment analysis);
- integrating all modules into the MindSpaces architecture.

6.2 Related work summary

In this section, we compile the available annotated data and the state-of-the-art analysis tools in the five targeted languages of MindSpaces: English, Spanish, Catalan, Greek and French.

6.2.1 Existing corpora

Corpora of annotated sentences are needed in order to train statistical analysers (e.g., part-of-speech taggers, lemmatizers, or syntactic parsers). For all languages, UPF is developing Universal Dependency (UD)-based tools.

Table 10 displays the main features of the reference UD corpora we use:

Table 10. English, Spanish, Catalan, Greek and French UD corpora

Name	Short description	Format	Size	License	Language
Universal Dependencies (UD) (Nivre, et al. 2016)	Manually revised version of open textual material from electronic journal articles, blogs, etc.	CoNLL-X[6]	~16,000 sentences ~150,000 tokens.	GNU GPL 3.0	English
UD-Spanish Ancora (Martínez Alonso and Zeman 2016)	Automatically converted from AnCora[7].	CoNLL-X uses 17 UPOS tags	17,680 sentences, 547,681 tokens and 549,570 syntactic words.	GNU GPL 3.0	Spanish
UD-Catalan Ancora (Martínez Alonso and Zeman 2016)	Automatically converted from AnCora.	CoNLL-X uses 17 UPOS tags	16,678 sentences, 530,766 tokens and 546,680 syntactic words.	GNU GPL 3.0	Catalan

[6] https://universaldependencies.org/format.html

[7] AnCora consists mainly of newspaper texts annotated at different levels of linguistic description: morphological (PoS and lemmas), syntactic (constituents and functions), and semantic (argument structures, thematic roles, semantic verb classes, named entities, and WordNet nominal senses). All resulting layers are independent of each other.

UD_Greek-GDT (Prokopidis and Papageorgiou 2017)	The Greek UD treebank is derived from the Greek Dependency Treebank (http://gdt.ilsp.gr), a resource developed and maintained by researchers at the Institute for Language and Speech Processing/Athena R.C. (http://www.ilsp.gr)	CoNLL-X	61,673 tokens 63,441 words 2,521 sentences	Creative Commons License Attribution-ShareAlike, CC BY-NC-SA 3.0	Greek
UD_French-GSD	The UD_French-GSD was converted in 2015 from the content head version of the universal dependency treebank v2.0		400,396 words 16,341 sentences	Creative Commons License Attribution-ShareAlike 4.0 International	French

6.2.2 Existing tools for syntactic and semantic parsing

A very large amount of NLP tools has been developed in recent years; most tools are language-agnostic and simply need to be trained on the resources of a desired language. One of the most widely used toolkits is Stanford CoreNLP (Manning, et al. 2014), which contains all the basic components needed in an NLP analysis pipeline: sentence splitting, tokenization, lemmatization, morphological tagging, coreference resolution, dependency parsing. Other popular toolkits are MATE Tools (Bohnet and Nivre 2012), developed at the university of Stuttgart, and Nlp4J[8]. We currently use components of these off-the-shelf toolkits, which we trained for our purposes, as shown in Table 11:

Table 11: Off-the-shelf tools used in the MindSpaces analysis pipeline

	English	**Spanish**	**Catalan**	**Greek**	**French**
Segmenter	Stanford CoreNLP v3.8.0	Stanford CoreNLP v3.8.0	Stanford CoreNLP v3.8.0	Stanford CoreNLP v3.8.0	Stanford CoreNLP v3.8.0

[8] https://emorynlp.github.io/nlp4j/

PoS Tagger	Stanford CoreNLP v3.8.0	Stanford CoreNLP v3.8.0	Stanford CoreNLP v3.8.0	Stanford CoreNLP v3.8.0	Stanford CoreNLP v3.8.0
Lemmatiser	Mate Tools v3.5	Mate Tools v3.5	Mate Tools v3.5	Mate Tools v3.5	Mate Tools v3.5
Morph Tagger	Mate Tools v3.5	Mate Tools v3.5	Mate Tools v3.5	Mate Tools v3.5	Mate Tools v3.5
Dependency parser	Nlp4J v1.1.3	Nlp4J v1.1.3	Nlp4J v1.1.3	Nlp4J v1.1.3	Nlp4J v1.1.3

6.2.3 **Existing tools for concept extraction**

There is a vast variety of different methods for information extraction and named entity recognition that are suitable for the concept extraction task. The main state-of-the-art models, algorithms, and tools are listed in Table 12. Their peculiarities and drawbacks are highlighted.

Table 12: Concept extraction tools

OLLIE	OLLIE – Open Language Learning for Information Extraction (Schmitz, et al. 2012) - is an open Information Extraction (IE) system for extracting relational tuples from text, without requiring a pre-specified vocabulary, by identifying relation phrases and associated arguments in arbitrary sentences. It outperforms its strong predecessors REVERB (Fader, Soderland and Etzioni 2011) and WOEparse (Wu and Weld 2010) by addressing their limitations – (1) they extract only relations that are mediated by verbs, and (2) they ignore context, thus extracting tuples that are not asserted as factual. First, OLLIE achieves high yield by extracting relations mediated by nouns, adjectives, and more. Second, a context-analysis step increases precision by including contextual information from the sentence in the extractions. OLLIE obtains 2.7 times the area under precision-yield curve (AUC) compared to REVERB and 1.9 times the AUC of WOEparse. Concept extraction is not the primary goal of OLLIE since it aims at detecting relations and not all concepts participate in relations. Therefore some concepts are missed causing a low recall.
AutoPhrase	AutoPhrase (Shang, et al. 2018) is a framework for automated phrase mining which leverages this large amount of high-quality phrases in an effective way and achieves better performance compared to limited human labelled phrases. It is based on positive-only distant training with random forest (Geurts, Ernst and Wehenkel 2006) for phrase classification. It also provides a PoS-guided phrasal segmentation model, which incorporates the shallow syntactic information in part-of-speech (PoS) tags to enhance the performance. The drawback of the approach is that it checks an exhaustive set of n-grams without straight restrictions on possible combinations of PoS-tags and therefore leaves a chance to outcome low-quality phrases in case they add a higher value to the overall score calculated with dynamic programming algorithms.

SpaCy NER	SpaCy (Honnibal and Montani 2017) features a fast statistical entity recognition system, which assigns labels to contiguous spans of tokens. The default model identifies a variety of named and numeric entities, including companies, locations, organisations, and products. As it was trained exclusively on named entities it might miss some real-world concepts that are not names.
AIDA	AIDA (Yosef, et al. 2011) is a framework and an online tool for entity detection and disambiguation. Given a natural-language text or a Web table, it maps mentions of ambiguous names onto canonical entities (e.g., individual people or places) registered in the YAGO2 knowledge base. It mostly focuses on capitalised named entities resulting at high precision with a comparably low recall.
FRED	FRED (Gangemi, et al. 2017) is a machine reader for the semantic web: its output is an RDF/OWL graph, whose design is based on frame semantics. Nevertheless, FRED's graphs are domain- and task-independent making the tool suitable to be used as a semantic middleware for domain- or task-specific applications. To serve this purpose, it is available both as a REST service and as a Python library. It detects a hierarchical set of relations between entities that sometimes leads to problems with the processing of long sentences with a large number of entities.
DBpedia Spotlight	DBpedia Spotlight (Daiber, et al. 2013) is a system for automatically annotating text documents with DBpedia URIs. DBpedia Spotlight allows users to configure the annotations to their specific needs through the DBpedia Ontology and quality measures such as prominence, topical pertinence, contextual ambiguity, and disambiguation confidence. DBpedia Spotlight is shared as open source and deployed as a Web Service freely available for public use. It heavily relies on large gazetteers built on top of entire Wikipedia for interconnecting the Web of Documents with the Web of Data. Integration of an exhaustive number of real-world entities makes this system one of the most competitive on the market.
Lample et al., 2016	Lample (Lample, et al. 2016) provides a state-of-the-art named entity recognition model that avoids heavily relying on traditional hand-crafted features and domain-specific knowledge. The model is a neural architecture based on bidirectional LSTMs and conditional random fields that relies on two sources of information about words: character-based word representations learned from the supervised corpus and unsupervised word representations learned from unannotated corpora. It obtains state-of-the-art performance in NER in four languages without resorting to any language-specific knowledge or resources such as gazetteers. As other named entity recognition tools it mostly focuses on capitalised words sometimes missing entities written in lower-case.
BERT NER	BERT (Devlin, et al. 2019) is a state-of-the-art language representation model, which stands for Bidirectional Encoder Representations from Transformers. Unlike recent language representation models, BERT is designed to pre-train deep bidirectional representations from unlabelled text by jointly conditioning on both left and right context in all layers. As a result, the pre-trained BERT model can be fine-tuned with just one additional output layer to create state-of-the-art models for a wide range of tasks including named entity recognition without substantial task-specific architecture modifications. At the same time, such universality of the model might lead to weaker results in comparison to models specially designed for a particular task.
Flair NER	Flair (Akbik, et al. 2019) is a library that provides state-of-the-art natural language processing models including a named entity recognition model. The core idea of

	the framework is to present a simple, unified interface for conceptually very different types of word and document embeddings. This effectively hides all embedding-specific engineering complexity and allows researchers to "mix and match" various embeddings with little effort. The framework also implements standard model training and hyperparameter selection routines, as well as a data fetching module that can download publicly available NLP datasets and convert them into data structures for quick set up of experiments. The NER model shows an improvement of the results over the best models for many standard datasets.

Related work analysis shows the necessity in developing new methods that would address drawbacks of the state-of-the-art tools. Following the well-established tendency in information extraction adopted for NER and extractive summarization, we envisage concept extraction as an attention-based sequence-to-sequence learning problem.

6.3 Processing Pipeline Definition

6.3.1 Text pre-processing

A common set of pre-processing operations are applied to input texts of any type (short Twitter messages, long articles of newspapers and blogs, transcripts of interviews). Pre-processing steps include:

- Segmentation: detection of sentence boundaries (if more than one sentence in the input);
- Lemmatisation: prediction of the base form of the words (painting VS paintings);
- PoS tagging: assignment of grammatical categories (painting = NOUN, paint = VERB)
- Morphological analysis: detection of morphological features (paints = paint+PRESENT+SINGULAR+3rdPERSON+INDICATIVE).

The tools used for the pre-processing are listed in Table 11.

6.3.2 Syntactic and semantic parsing

The current NLP analysis pipeline outputs two different types of structures, which correspond to three different levels of abstraction of the linguistic description:

- SSynt: surface-syntactic structures (SSyntSs), i.e., language-specific syntactic trees with fine-grained relations over all the words of a sentence;
- DSynt: deep-syntactic, or *shallow-semantic*, structures (DSyntSs), i.e., language-independent syntactic trees with coarse-grained relations over the meaning-bearing units of a sentence.

This stratified view is strongly influenced by the Meaning-Text Theory -MTT (Mel'čuk 1988). The MTT model supports fine-grained annotation at the three main levels of the linguistic description of written language: semantics, syntax and morphology, while facilitating a

coherent transition between them via intermediate levels of deep-syntax and deep-morphology. At each level, a clearly defined type of linguistic phenomena is described in terms of distinct dependency structures.

In the framework of MindSpaces, UPF is using a Universal Dependency-based pipeline, which uses similar approaches and tag sets across languages. Universal Dependencies is a generic framework for cross-lingual syntactico-semantic annotation that has been applied to over 80 languages so far, for a total of over 140 different treebanks[9]. Most treebanks have been obtained through automatic conversions of other treebanks, themselves in general obtained via automatic annotation. The resulting annotations are known to lack consistency and quality, but they have the advantage to provide a framework that reduces the differences across different languages. In MindSpaces, we intend to test the usability of Universal Dependencies as intermediate representations for multilingual relation extraction.

For surface-syntactic parsing, we train the off-the-shelf Nlp4J parser on the freely available UD corpora of the MindSpaces languages (English, Spanish, Catalan, Greek and French). The resulting surface structures are syntactic trees with lemmas, part-of-speech tags, morphological and dependency information under the form of grammatical functions such as *subject*, *object*, *adverbial*, etc.

The deep structures in this configuration consist of predicate-argument structures obtained through the application of graph-transduction grammars to the UD surface-syntactic structures. The deep and surface structures are aligned node to node. In the deep structures, we aim at removing all the information that is language-specific and oriented towards syntax:

- Determiners and auxiliaries are replaced (when needed) by attribute/value pairs, as, e.g., Definiteness, Aspect, and Mood:
  - Auxiliaries: *was painted-> paint*;
  - Determiners: *the paintings-> painting*;
- Functional prepositions and conjunctions that can be inferred from other lexical units or from the syntactic structure are removed;
  - *painted by X-> painted X*
- Edge labels are generalised into predicate argument (semantics-oriented) labels in the PropBank/NomBank fashion:
  - *subject*(*painted, by X)-> FirstArgument(paint, X)*

The UD-based pipeline doesn't make any use of lexical resources at this point; the predicate-argument relations are derived using syntactic cues only. The deep input is a compromise between (i) correctness and (ii) adequacy in a generation setup. Indeed, the conversion of the UD structures into predicate-argument structures depends not only on the mapping process, but also on the availability of the information in the original annotation.

[9] http://universaldependencies.org/

Table 13 shows different labels that the UD-based graph-transduction grammars currently produce.

Table 13: Semantic labels in the output of the UD-based pipeline

Semantic label	**Type**	**Description**	**Example**
A1/A1INV	Core	1st argument of a predicate	paint-> an artist
A2/A2INV	Core	2nd argument of a predicate	paint-> a painting
A3/A3INV	Core	3rd argument of a predicate	donate-> to a museum
A4, A5, A6	Core	4th to 6th arguments	*Very uncommon*
AM	Non-Core	None of governor or dependent are argument of the other	exhibited-> in a gallery
LIST	Coordinative	List of elements	painted-> and-> exhibited
NAME	Lexical	Part of a name	Louvre-> Museum
DEP	UKN	Undefined dependent	N/A

The following phenomena should be highlighted:

- **Alignment between surface and deep nodes**: On the deep nodes, we use one or more feature ids with attributed as suffix the line number of the corresponding surface nodes: on a deep node, id1=4|id2=15 means that this deep node is aligned with the surface nodes on the lines 4 and 15 of the corresponding surface structure. Only elements triggered by other elements (as opposed to be triggered by the structure of the sentence) are aligned with deep nodes. That is, a subcategorised preposition is aligned with a deep node, while a void copula or an expletive subject are not.

- **Core relations**: Each defined core relation is unique for each predicate: there cannot be two arguments with the same slot for one predicate. If a predicate has an A2 dependent, it cannot have another A2 dependent, and it cannot be A2INV of another predicate.

- **Auxiliaries:** Auxiliaries are mapped to the universal feature "Aspect"[10].

- **Conjunctions/prepositions:** The prepositions and conjunctions maintained in the deep representation can be found under an A2INV dependency. A dependency path Gov-AM-> Dep-A2INV-> Prep is equivalent to a predicate (the conjunction/preposition) with 2 arguments: Gov <-A1-Prep-A2-> Dep.

- **Modals:** They are mapped to the universal feature "Mood".

- **Pronouns:**
    - Relative: only subject and object relative pronouns directly linked to the main relative verb are removed from the deep structure.
    - Subject: a dummy pronoun node for subject is added if an originally finite verb has no first argument and no available argument to build a passive; for a pro-drop language such as Spanish, a dummy pronoun is added if the first argument is missing.

- **Punctuations:** Only the final punctuations are encoded in the deep representations: the main node of a sentence indicates if the latter is declarative, interrogative, exclamative, suspensive, or if it is involved in a parataxis, with the feature "clause_type".

Our graph-transduction grammars are rules that apply to a subgraph of the input structure and produce a part of the output structure. During the application of the rules, both the input structure (covered by the left side of the rule) and the current state of the output structure at the moment of application of a rule (i.e., the right side of the rule) are available as context. The output structure in one transduction is built incrementally: the rules are all evaluated, the ones that match a part of the input graph are applied, and a first piece of the output graph is built; then the rules are evaluated again, this time with the right-side context as well, and another part of the output graph is built; and so on. The transduction is over when no rule is left that matches the combination of the left-side and the right-side.

Consider, for illustration, a sample rule from the SSynt-DSynt mapping in Figure 60. This rule, in which we can see the left-side and the right-side fields, removes auxiliaries, such as *was* in *was painted,* and gets the verbal finiteness up to the full verbs. The right-side context is indicated by the prefix *rc:* before a variable or a correspondence; in our example, it means that the rule looks for the *rc: -compound tenses* in the current state of the output structure, and builds the elements (full verbs) with no dependent, but with the new features *VerbForm* and *original_VerbForm*, which store the left-side *VerbForm* of the dependent and the full verb, respectively. A similar rule would apply to *bring* and *can* in Figure 61; as a result of the application of this rule, only *bring* is left in Figure 62, which has a correspondence with both *bring* and *can* from Figure 61.

[10] http://universaldependencies.org/u/feat/index.html

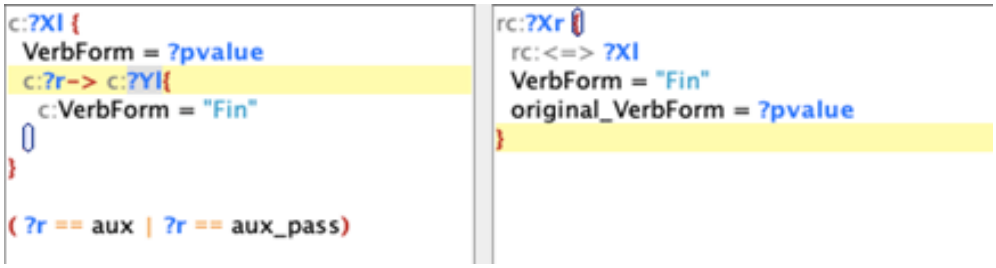


Figure 60: A sample graph-transduction rule; *?* indicates a variable; *?Xl{}* is a node, ?s-> is a relation, *a=?b* is an attribute/value pair.

Table 14: Graph-transduction rules for UD-based deep parsing. *Includes rules that simply copy node features (~40 per grammar)sums up the current state of the graph-transduction grammars and rules for the mapping between surface-syntactic structures and UD-based semantic structures that we have implemented for English, Spanish and Catalan during the first half of the project.

Table 14: Graph-transduction rules for UD-based deep parsing. *Includes rules that simply copy node features (~40 per grammar)

Grammars	**# rules***	**Description**
Normalization	62	Normalize dependencies
Pre-processing	85	Identify nodes to be removed Identify verbal finiteness and tense
SSynt-Sem	128	Remove idiosyncratic nodes Establish correspondences with surface nodes Predict predicate-argument dependency labels Replace determiners, modality and aspect markers by attribute-value feature structures Identify duplicated core dependency labels below one predicate
Post-processing	75	Replace duplicated argument relations by best educated guess Identify remaining duplicated core dependency labels (for posterior debugging)

Figure 61 and Figure 62 respectively show a syntactic structure as parsed by the integrated parser and the semantic structure produced by the graph-transduction grammars for the

sentence *When we see a work of art it can only bring well-being.* The pronouns *we* and *it* are correctly identified as the first arguments (A1) of *we* and *bring respectively*, the *obj* in the syntactic structure are identified as the second arguments (A2), and the relations between the subordinate clause (*when we see a work of art*) and *bring* and between the nominal modifier (*art*) and *work* are correctly identified as non-core (AM). The relations with the suffix *INV* (e.g. between *of* and *art*) indicate an inverted core relation between the two elements; their purpose is to maintain a tree format (in which every node has at most one governor), easier to process, as opposed to a graph format (in which a node can have several governors).

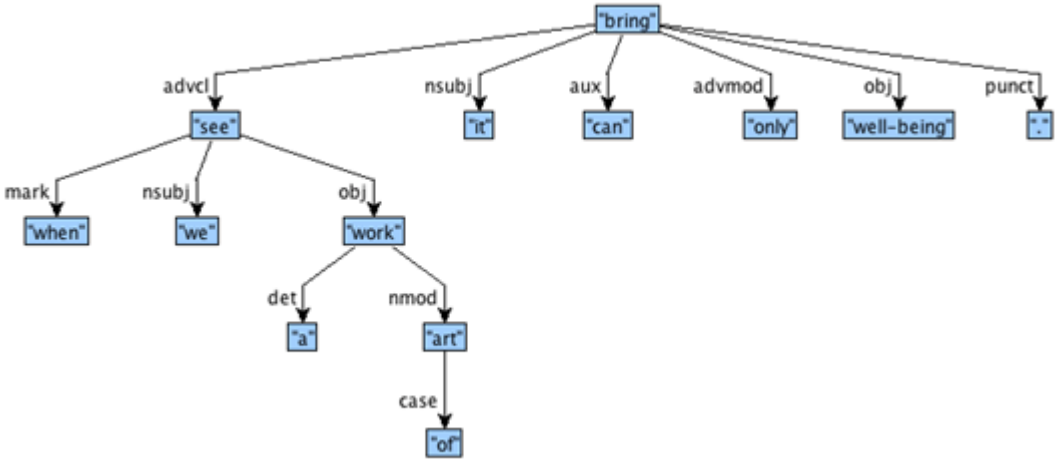


Figure 61: Surface-syntactic UD-Structure: *When we see a work of art it can only bring well-being.*

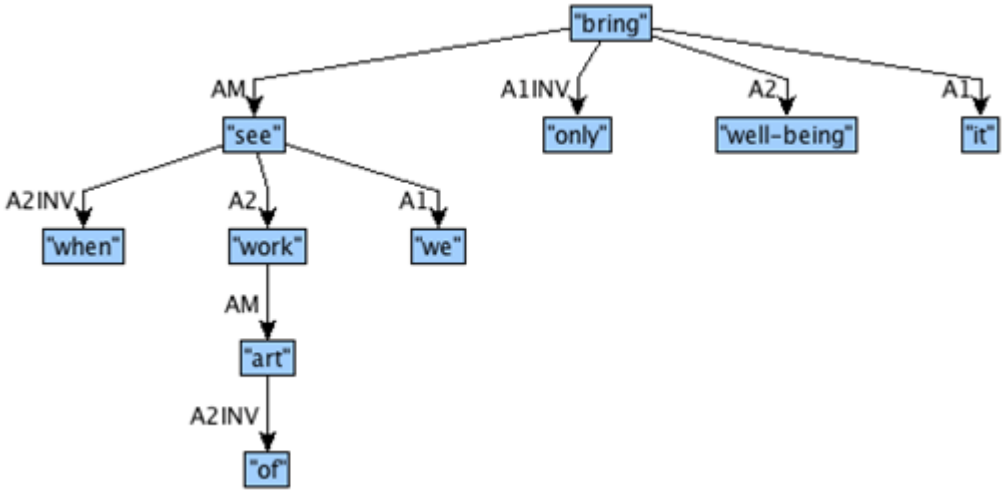


Figure 62: UD-based predicate-argument structure: *When we see a work of art it can only bring well-being.*

6.3.3 **Preliminary evaluation**

In Table 15, we provide an evaluation of the UD-based syntactic dependency parsing, using the most commonly used evaluation metrics in dependency parsing: *labelled* and *unlabelled attachment scores* (LAS/UAS). For this, we use the official UD test sets as provided in the

CoNLL 2017 shared task[11], because there is no annotated reference dataset pertinent to the MindSpaces domain. A formal evaluation of the deep-syntactic structures has not been carried out at this point.

Table 15: Results of the evaluation of the UD-based joint parsing

	UAS	**LAS**
English	0.87	0.85
Spanish	0.89	0.87
Catalan	0.91	0.88
Greek	0.87	0.82

6.4 Concept extraction

Concept extraction is one of the key tasks in language understanding that is applied in MindSpaces for discourse analysis. In order to retrieve concepts from general discourse textual material, we propose a generic open-domain OOV-oriented extractive model that is based on distant supervision of a pointer-generator network leveraging bidirectional long short-term memory (LSTM) units and a copy mechanism.

6.4.1 Task definition

In knowledge discovery and representation, the notion of concept is most often used to refer to sense, i.e., “abstract entity” or “abstract object” in the Fregean dichotomy of sense vs. reference (Frege 1892). In Natural Language Processing (NLP), the task of Concept Extraction (CE) deals with the identification of the language side of the concept coin, i.e., Frege's reference. Halliday (Halliday 2013) offers a syntactic interpretation of reference. In his terminology, it is a “classifying nominal group". For instance, *renewable energy* or *nuclear energy* are classifying nominal groups: they denote a class (or type) of energy, while, e.g., *cheap energy* or *affordable energy* are not: they do not typify, but rather qualify energy (and are thus “qualifying nominal groups").

We aim to extract from unstructured textual material classifying nominal groups, including individual nouns and nominal groups that correspond to an atomic entity (such as, e.g., *energy*) or widely used class in the real world.

[11] http://universaldependencies.org/conll17/

6.4.2 Description of the model

As a basis of our model, we use the pointer-generator network proposed in (See, Liu and Manning 2017) that aids creation of summaries with accurate reproduction of information. In each generation step *t*, the *pointer* allows for copying words w_i from the source sequence to the target sequence using distribution of attention layer a^t, while the *generator* samples tokens from the learned vocabulary distribution P_{vocab}, conditioned by a context vector h_t^* produced by the same attention layer which is built based on hidden states h_i of an encoder and states s_t of a decoder (in each case, a bidirectional LSTM (Graves and Schmidhuber 2005)). In addition, coverage mechanism is applied to modify at using a coverage vector c_t to avoid undesirable repetitions in the output sequence. Specifically, to produce a word *w*, the above-mentioned distributions are combined into a single final probability distribution being weighted using the *generation probability* $p_{gen} \in [0,1]$:

$$P(w) = p_{gen} \cdot P_{vocab}(w) + (1 - p_{gen}) \sum_{i:w_i=w} a_i^t$$

where $P_{vocab}(w)$ is the vocabulary distribution, which is zero if *w* is an out-of-vocabulary (OOV) word; a^t is the attention distribution; w_i - tokens of the input sequence; $\sum_{i:w_i=w} a_i^t$ is zero if *w* does not appear in the source sequence. According to (See, Liu and Manning 2017), individual vectors, distributions, and probability p_{gen} are defined as follows:

$$c^t = \sum_{t'=0}^{t-1} a^{t'},$$
$$e_i^t = v^T tanh(W_h h_i + W_s s_t + w_c c_i^t + b_{attn}),$$
$$a^t = softmax(e^t),$$
$$h_t^* = \sum_i a_i^t h_i,$$
$$P_{vocab} = softmax(V'(V[s_t, h_t^*] + b) + b'),$$
$$p_{gen} = \sigma(w_{h^*}^T h_t^* + w_s^T s_t + w_x^T x_t + b_{ptr}),$$

where v, W_h, W_s, w_c, b_{attn}, V, V', b, b', w_{h^*}, w_s, w_x, b_{ptr} are learnable parameters, *T* stands for the transpose of a vector, x_t is the decoder input, and σ is the sigmoid function.

To adapt this basic model to the task of CE, we applied several modifications to it (cf., Figure 63): (i) following (Gu, et al. 2016), we use separate distributions for copy attention and general attention, instead of one for both; (ii) experiments have shown that encoders and decoders with several LSTM layers perform better than with a single layer, such that we work with multiple layer LSTMs; how many is determined using a development dataset; (iii) we adapt the forms of input and target sequences to the specifics of the task of CE. The input is comprised of tokens and their part-of-speech (PoS) tags (e.g., “The DT President NN is VBZ elected VBD by IN a DT direct JJ vote NN”). The target sequence concatenates concepts in the order they appear in the text and separates them by a token “*" especially introduced to partition the output (e.g., “President * direct vote”).

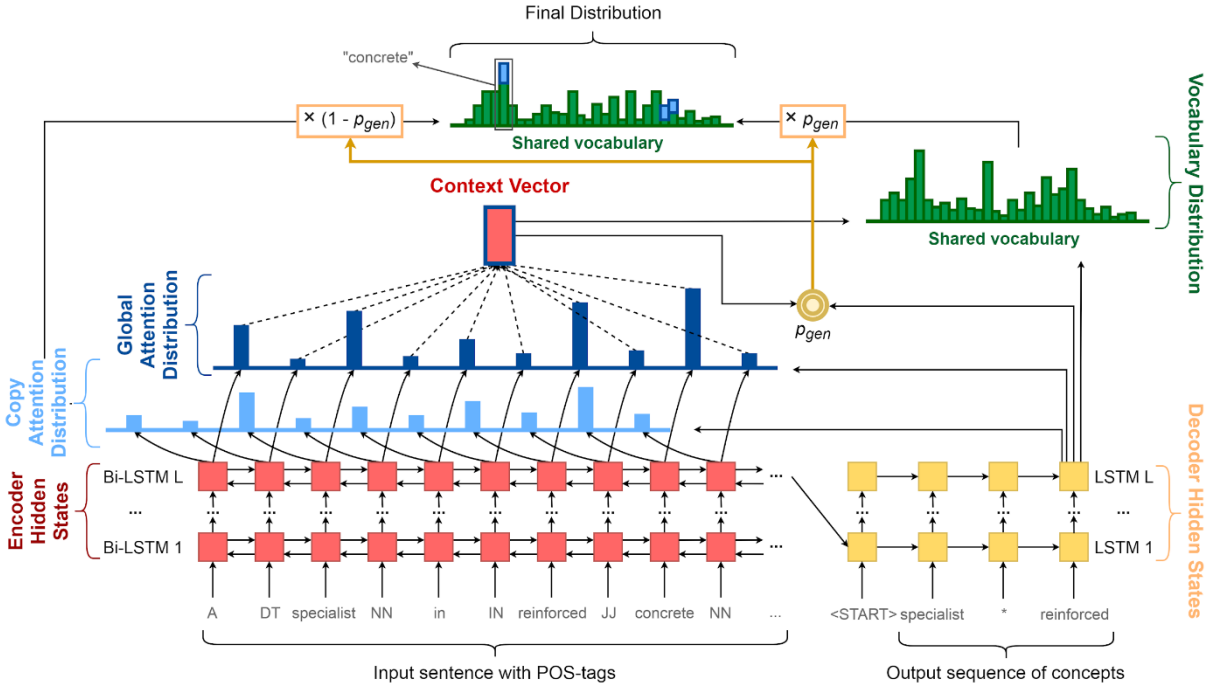


Figure 63: The neural architecture for concept extraction

This model is naturally applicable to the task of CE since it facilitates the selection and transfer of subsequences of tokens (= concepts) from a given source sequence of tokens (= text input) to the target sequence (= partitioned sequence of concepts). The pointer mechanism implies the ability to cope with OOV words, which is crucial for universal CE, while the generator implies the ability to adjust vocabulary distribution for selecting the next word (which might be a termination token "*") based on a given context vector, which allows us to implicitly take into account the domain specifics and linguistic features that facilitate the task of CE. Furthermore, the updating of vocabulary distribution adds the possibility to vanish or strengthen the copy effect and thus learn to distinguish concepts with outer modiers (such as, e.g., "hot *air*", "[fully] crewed *aircraft*", "reinforced *group*") from multiword concepts (such as, e.g., "*hot air balloon*", "*unmanned aerial vehicle*", "*reinforced concrete*").

6.4.3 Compilation of the training corpus

We automatically create a (noisy) training corpus using two various annotators over a large unlabelled corpus of Wikipedia articles: DBpedia Spotlight (Daiber, et al. 2013) with the value of its confidence coefficient that gains the highest recall and our own algorithm that uses a number of rules and heuristics. Our labelling is based on the sentence-wise analysis of statistical and linguistic features of sequences of tokens. First, named entities and multiple token concepts and then single token concepts are identified. The algorithm covers the following tasks:

Application of a statistical NER model. A significant number of concepts in Wikipedia are capitalized terms, which can be captured by statistical named entity recognizers (NER); see the Related Work section above. Therefore, at first, SpaCy's state-of-the-art NER model (Honnibal and Montani 2017) is applied with a successive elimination of used tokens for

further processing. The next steps are applied then separately to fragments of texts located between the identified NEs.

Selection of n-grams as fragments of NP chunks that can form part of multiple token concepts. For this task, we formed the PoS-patterns based on Penn Treebank tagset[12], which were inherited from the patterns for multiword expression detection introduced in (Cordeiro, Ramisch and Villavicencio 2016) and expanded here resulting in the following set: *P* = {N_N, J_N, V_N, N_J, J_J, V_J, N_of_N, N_of_DT_N, N_of_J, N_of_DT_J, N_of_V, N_of_DT_V, CD_N, CD_J}, where N stands for "noun", i.e., NN|NNS|NNP|NNPS, J stands for "adjective", i.e., JJ|JJR|JJS, V - "verb" but limited to VBD|VBG|VN, CD - "cardinal number", DT - "determiner", and "of" is an exact pronoun. Each pattern matches an n-gram with two open-class lexical items and at most two auxiliary tokens between them.

Assessment of the distinctiveness of each selected n-gram. The distinctiveness of selected *n*-grams is assessed using word co-occurrences from the Google Books dataset[13]. Let us assume a given an n-gram $T_1A_1A_2T_2 \in c_k$, where T_1 and T_2 are open class lexical items and A_1 and A_2 are optional auxiliary tokens, and c_k is a set of all n-grams of a particular kind of pattern $p_k \in P$. We use $T_1A_1A_2T_2$ as a point of a function that passes through normalized document frequencies of a set of similar *n*-grams $\cup_{c_k} T_1A_1A_2T_j$, arrayed in ascending order, to find a tangential angle at this point. Similarly, a tangential angle at the point $T_1A_1A_2T_2$ on a curve of ordered frequencies of n-grams $\cup_{c_k} T_hA_1A_2T_2$ is determined. We leverage these angles to check how prominent an n-gram, i.e., to what extent it differs from its neighbours by overall usage. In case an n-gram is located among equally prominent n-grams with a tangential angle close to 0, we do not consider it as a potential part of a concept since it does not show a notable distinctiveness inherent in concepts, especially in common idiosyncratic concepts. The thresholds Q_{min1} and Q_{min2} for a minimum allowed angles of a slope among the sets $\cup_{c_k} T_1A_1A_2T_j$ and $\cup_{c_k} T_hA_1A_2T_2$ are to be predefined. The Q_{min1} concerns the maximum value of two angles and Q_{min2} – the minimum value of them.

Combination of intersected highly distinctive parts as concepts. We combine those distinctive n-grams that share common tokens and iteratively drop the last token in each group if it is not a noun, in order to end up with complete NP candidate concepts (e.g., "value of the played card" is a potential concept corresponding to the patterns {N_of_DT_V; V_N}). Some single-word concepts already might appear at this point.

Recovery of missed single-word concepts. To enrich the set of candidate concepts, we consider all unused nouns and numbers in a text as single-word concept candidates.

The obtained training corpus contains moderate amount of noise: the proposed annotation algorithm outperforms some baselines and might be used for CE by itself (see the model $DSA_{(60;0)}$ in Table 5).

[12] https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html

[13] https://books.google.com/ngrams/

6.4.4 Realization

For training, token sequences are taken from annotated sentences with a sliding overlapping window of a fixed maximum length, which is minimally expanded if needed in order not to deal with incomplete concepts at the borders. The trained model is applied to unseen sentences, which are also split into sequences of tokens with an overlapping window of the same size, without any expansion. Finally, the corresponding mentions in the plain text are determined since the output format does not include offsets. In particular, following (Hasibi, Balog and Bratsberg 2015), we find all possible matches for all detected concepts and then successively select non-nested concepts from the beginning to the end of the sentence, giving priority to the longest, in case of a multiple choice.

For the realization of the model, we use the implementation of See et al.'s pointer-generator model in the OpenNMT toolkit (Klein, et al. 2018), which allows for the adaptation of the model to the task of CE along the lines described in Section 3.1 above. Instead of the attention mechanism used in (See, Liu and Manning 2017), we use the default OpenNMT attention (Luong, Pham and Manning 2015) since it showed to perform better. The model has 512-dimensional hidden states and 256-dimensional word embeddings shared between encoder and decoder. We use a vocabulary of 50k words as we rely mostly on a copying mechanism which uses dynamic vocabulary made up of words from the current source sequence. We train using the Stochastic Gradient Descent on a single GeForce GTX 1080 Ti GPU with a batch size of 64. We trained the CE-adapted pointer-generator networks of two and three bi-LSTM layers with 20K and 100K training steps and selected the best performing models using the development set. For languages, where both Google Books dataset and DBpedia Spotlight were available, we combined models trained separately with different annotation schemes (see above) into an ensemble. For poorly resourced languages like Catalan, links to other pages within a text of the Wikipedia article were used for the training as ground truth concepts.

6.4.5 Preliminary evaluation

Two types of evaluations were carried out: a generic one to compare our model with the state-of-the-art approaches, and a domain-specific, where we used materials of textual collections created within the project.

The performance is measured in terms of precision, recall, and F1-score, aiming at high recall, first of all. Since there are no negative examples in ground truth annotation, we treated only the detected concepts that partially overlapped the ground truth concepts as false positives. Concepts that have the same spans as the ground truth concepts are counted as true positives, and missed ground truth concepts as false negatives. This meets our goal to detect the exact match. It also allows us to penalize brute force high-recall algorithms that produce a large number of candidates, which are of limited use in real-world applications.

An annotated snapshot of Wikipedia (Schenkel, Suchanek and Kasneci 2007) was used for the generic evaluation. Results are provided in Table 16. "DSA" stands for distant supervision annotation, in particular, "DSA_{DICT}" when annotation was obtained using DBpedia Spotlight,

i.e., a dictionary lookup, and "$DSA_{(60;0)}$" when the proposed token-cooccurrence frequency-based method was applied (cf. the compilation of the training corpus). The values in parentheses correspond to Q_{min1} and Q_{min2}, which gave the best scores on the development set; $PG_{(2L;18K)}$ and $PG_{(3L;80K)}$ stand for pointer-generator networks with parameters shown in parentheses.

Table 16: Concept extraction evaluation on generic open-domain texts

Setup	Model	Precision	Recall	F1-score
	FLAIR (Akbik, et al. 2019)	**0.79**	0.59	0.67
	AutoPhrase+ DBLP (Shang, et al. 2018)	0.4	0.43	0.41
	AutoPhrase+ WIKI (Shang, et al. 2018)	0.43	0.49	0.46
	NER Tagger (Lample, et al. 2016)	0.77	0.58	0.66
	WAT (Piccinno and Ferragina 2014)	0.68	0.42	0.52
	$Spotlight_{0.1}$ (Daiber, et al. 2013)	0.69	0.73	0.71
	OLLIE (Schmitz, et al. 2012)	0.44	0.18	0.26
	AIDA (Yosef, et al. 2011)	0.77	0.45	0.57
(A)	$DSA_{(60;0)}$	0.65	0.72	0.68
(B)	$PG_{(3L;80K)}(DSA_{DICT})$	0.68	0.72	0.7
(C)	$PG_{(2L;18K)}(DSA_{(60;0)})$	0.68	0.76	0.72
(D)	(B) + (C)	0.72	**0.8**	**0.76**

Furthermore, we have evaluated the concept extraction component using test datasets for three languages of MindSpaces, namely, English, Spanish, and Catalan. These datasets were compiled from the set of tweets in Spanish and Catalan collected by the Crawler component within PUC1, and from the texts of design news blogs collected by the Scraper component within PUC2.

A set of sentences was manually annotated for each language resulting in more than 300 concepts in 50 to 400 sentences per language depending on the collection.

Since dictionary-based approaches are very competitive in concept extraction (Daiber, et al. 2013); (Piccinno and Ferragina 2014), we designed *two baseline algorithms* on top of a large scale lexicon that was obtained by the extraction of links from the whole dump of Wikipedia separately for each language. The first algorithm, that aims to guarantee high recall, consists in detecting all the entries from the lexicon in a given text allowing for overlap of the textual spans corresponding to concepts. The second algorithm, that is expected to increase the precision while still keeping recall at a high rate, consists in selecting the longest non-overlapping spans detected by the first algorithm in a successive order from the beginning to the end of the sentence.

Results of evaluation are presented in

Table 17 to Table 19.

Table 17: Concept extraction evaluation on texts from the PUC1 collection in Spanish

Algorithm	Precision	Recall	F1-score
Baseline1	0.48	0.82	0.61
Baseline2	0.53	0.69	0.60
Our algorithm	0.68	0.71	**0.70**

Table 18: Concept extraction evaluation on texts from the PUC1 collection in Catalan

Algorithm	Precision	Recall	F1-score
Baseline1	0.51	0.67	0.58
Baseline2	0.56	0.66	0.60
Our algorithm	0.64	0.73	**0.70**

Table 19: Concept extraction evaluation on texts from the PUC2 collection in English

Algorithm	Precision	Recall	F1-score
Baseline1	0.53	0.84	0.65
Baseline2	0.59	0.62	0.60
Our algorithm	0.67	0.69	**0.68**

The proposed algorithm significantly outperforms the baselines at all the datasets and reaches the equal F1-score of 0.7 for all languages. As expected, the first baseline provided the highest values of recall (at least for two languages) but as it generated an overly large number of candidates, the value of precision is very low. Manual analysis of the 30% of test examples, where concepts were not detected by our algorithm, showed that in many cases the ground truth spans were not matched precisely but the head words were identified correctly. This means that most information is preserved even if not the exact meaning is captured in some cases.

6.5 Semantic parsing

6.5.1 Generalized representation of predicate-argument structures

Capturing the underlying semantics and analysing the sentiments of citizens' social media contributions will allow for the contextual interpretation, by means of reasoning, within the overall frame of reference, and for the inference of general opinions that are useful to users.

For capturing the semantics, we currently use Predicate-argument (PredArg) structures, which are representations with abstract semantic role labels that also capture the underlying argument structure of predicative elements (which is not made explicit in syntax). To obtain these PredArg structures, we run another set of graph-transducers on the output of the DSynt parser.

Figure 64 shows the representation at the PredArg level of the sentence *When we see a work of art it can only bring well-being.* This layer of representation is very similar to the deep-syntactic / shallow semantic presented in Section 6.3.2 . The main differences are that: (i) lexical units are tagged according to an existing lexico-semantic resource, namely VerbNet, as e.g. *bring* and *see* in Figure 5, which are assigned the VerbNet classes 11.3 and 30.1, respectively (the current system is limited to choose the first meaning for each word); (ii) the suffix *INV* is removed and, for instance, A1INV, A2INV, as in Figure 3, become A1, A2, (iii) the prepositions, which in the deep representation were found under an A2INV dependency (Gov-AM-> Dep-A2INV-> Prep), are represented as predicates with 2 arguments: Gov <-A1-Prep-A2-> Dep, and (iv) the non-core relation between subordinate clauses and main verbs (AM) becomes *Elaboration.* In predicate-argument relations we also aim at removing support verbs. For the time being, this is restricted to light be-constructions, that is, constructions in which the second argument of *be* in the DSyntS is a predicate P that can have a first argument and that does not have a first argument in the structure. In this case, the first argument of the light *be* becomes the first argument of P in the PredArg representation; for instance, a structure like *painting <-I be II-> beautiful* is annotated as *beautiful A1-> painting*.

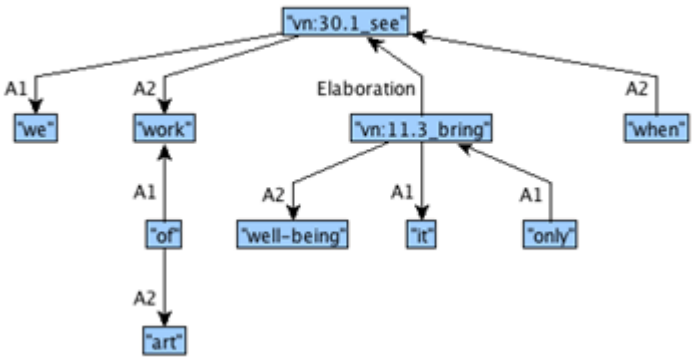


Figure 64: PredArg structure corresponding to the sentence *When we see a work of art it can only bring well-being*.

6.5.2 **Linking against external resources**

This section lists the tools used within the project for entity disambiguation via entity linking (EL) to external resources for the task of enriching linguistic representations with further semantic information. The task is defined in terms of deep syntactic or predicate-argument

relations holding between pairs of references to lexical and knowledge resources such as DBpedia[14], BabelNet[15], PropBank[16], NomBank[17], VerbNet[18] , WordNet[19], and FrameNet[20].

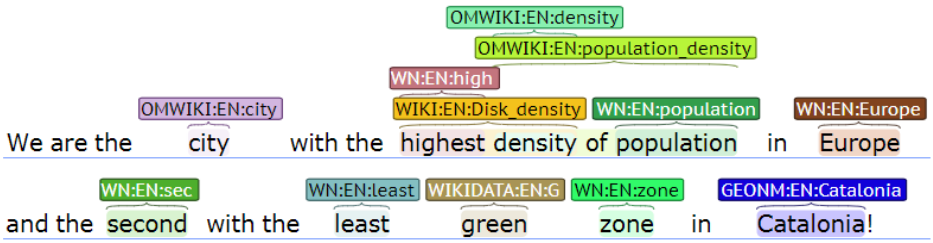


Figure 65: Example of entity linking results

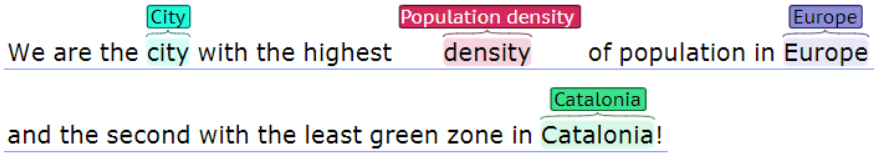


Figure 66: Results of entity linking with DBpedia Spotlight for English

The linguistic representation resulting from analysing the sentence “We are the city with the highest density of population in Europe and the second with the least green zone in Catalonia!” (in English) is illustrated in Figure 65 and Figure 66. NEs such as “Europe”, and “Catalonia” are detected and linked to their correct entries in WordNet, GeoNames, and DBpedia. Concepts like “city”, “density” and “population”, are also disambiguated to the correct entries in the aforementioned resources. Concepts from similar sentences in Spanish and Catalan (Figure 67 and Figure 68) are linked to external resources the same way. Spanish name “*Hospitalet de Llobregat*” and its Catalan version “*L’Hospitalet de Llobregat*” are disambiguated to the single entry in DBpedia “*Hospitalet de Llobregat*”.

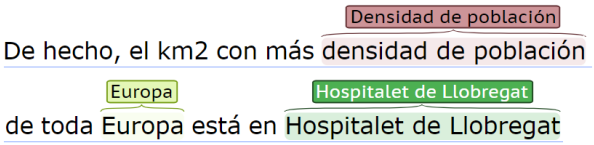
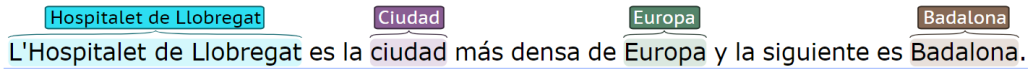


Figure 67: Results of entity linking with DBpedia Spotlight for Spanish



[14] https://wiki.dbpedia.org/

[15] https://babelnet.org/

[16] https://propbank.github.io/

[17] https://nlp.cs.nyu.edu/meyers/NomBank.html

[18] https://verbs.colorado.edu/~mpalmer/projects/verbnet.html

[19] https://wordnet.princeton.edu/

[20] https://framenet.icsi.berkeley.edu/fndrupal/

Figure 68: Results of entity linking with DBpedia Spotlight for Catalan

Let us consider another example of entity linking for Spanish, precisely, for the sentence “La arquitecta Pati Baztán cierra el año de arte urbano en Hospitalet de Llobregat” (Figure 69).

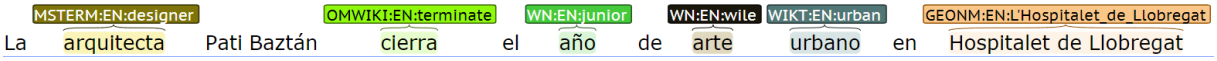


Figure 69: Example of entity linking results for Spanish

Figure 69 shows that the words “arquitecta” and “cierra” were linked to WordNet and Wikipedia and the corresponding universal labels “designer” and “terminate” were assigned. Deep dependency-based parsing is responsible for finding the correct senses for predicative words like “cierra” in PropBank, NomBank, VerbNet and FrameNet.

6.6 Sentiment analysis

Sentiment analysis is one of the key components in textual analysis for MindSpaces. As far as the task consists in the integration of existing sentiment analysis tools and their adaptation to the MindSpaces languages and use cases, we selected the most suitable among the publicly available models. We consider them in the following section.

6.6.1 Adaptation of existing tools

Emotion recognition

The recent public competition on emotion detection in text “SemEval-2019 Task 3 - EmoContext: Contextual Emotion Detection in Text“ (Chatterjee, et al. 2019) attracted many participants and a lot of state-of-the-art neural architectures were tested. In this task, given a textual dialogue i.e. an utterance along with two previous turns of context, the goal was to infer the underlying emotion of the utterance by choosing from four emotion classes - Happy, Sad, Angry and Others. A training data set of 30160 dialogues, and two evaluation data sets, Test1 and Test2, containing 2755 and 5509 dialogues respectively were released to the participants. The highest ranked submission achieved 79.59 micro-averaged F1 score.

According to the organizers of the competition, Bi-directional LSTM was the most common choice of neural architecture used, a good number of teams used the “Ekphrasis” package (Baziotis, Pelekis and Doulkeridis 2017) for tokenization, word normalization, word segmentation, and spell correction, and transfer learning was a popular choice among top teams. We selected the EmoSense model (Smetanin 2019) that possessed enumerated characteristics, achieved the score close to the top models, and at the same time was provided with the clear modifiable code[21].

The architecture of the chosen neural network (Figure 70) consists of the embedding unit and two bidirectional LSTM units (dim = 64). At the first step, each user utterance is fed into a corresponding bidirectional LSTM unit using pre-trained word embeddings. Next, feature maps are concatenated in a flatten feature vector and then passed to a fully connected

[21] https://github.com/sismetanin/emosense-semeval2019-task3-emocontext

hidden layer (dim = 30), which analyses interactions between obtained vectors. Finally, these features proceed through the output layer with the SoftMax activation function to predict a final class label. To reduce overfitting, regularization layers with Gaussian noise are used after the embedding layer, dropout layers (Srivastava, et al. 2014) are positioned at each LSTM unit (p = 0.2) and before the hidden fully connected layer (p = 0.1).

The important part of the approach is a set of pre-processing steps specific for the texts of dialogues that are also highly relevant to the texts of MindSpaces:

- URLs, emails, the date and time, usernames, percentage, currencies and numbers are replaced with the special corresponding tags;
- Repeated, censored, elongated, and capitalized terms are annotated with the special corresponding tags;
- Elongated words are automatically corrected based on built-in word statistics corpus;
- Hashtags and contractions unpacking (i.e., word segmentation) is performed based on built-in word statistics corpus;
- A manually created dictionary for replacing terms extracted from the text is used in order to reduce a variety of emotions;
- "Ekphrasis" package is used to identify most emojis, emoticons and complicated expressions such as censored, emphasized and elongated words as well as dates, times, currencies and acronyms.

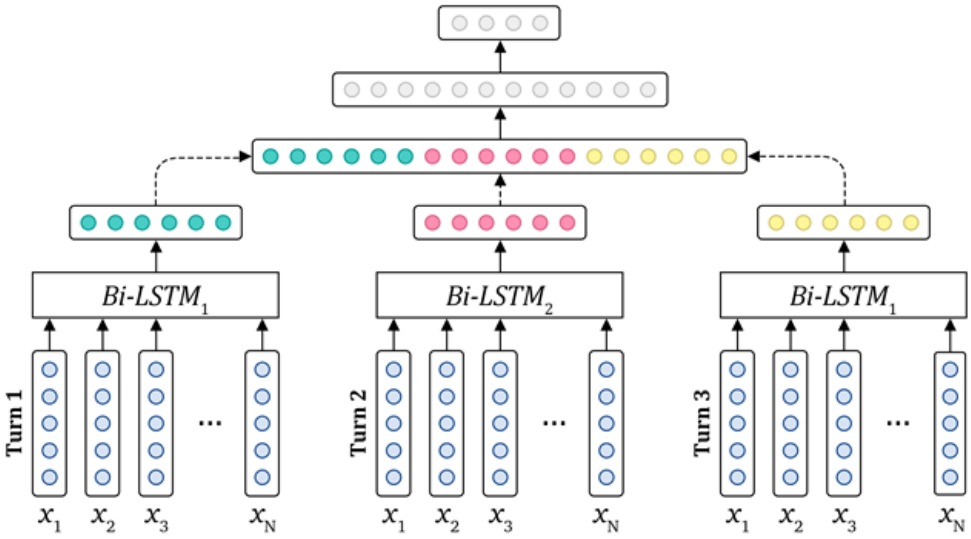


Figure 70: Architecture of neural network for emotion recognition

DataStories pre-trained word vectors (Baziotis, Pelekis and Doulkeridis 2017) are used as a basis. They were additionally fine-tuned on the automatically collected emotional dataset the following way: the embeddings layer was frozen for the first training epoch in order to avoid significant changes in the embeddings weights, and then it was unfrozen for the next 5 epochs. After the training stage, the fine-tuned embeddings were further used within supervised training phases with the SemEval-2019 EmoContext dataset.

On the final test dataset, the model achieved 72.59% micro-average F1-score for emotional classes. This is well above the official baseline released by task organizers, which was 58.68%.

We repurpose the pretrained model for individual short text classification by using only the latter bidirectional LSTM unit for the input text as it corresponded to the last turn in the dialogues where emotions were supposed to be found. Experiments show that the elimination of the context-related inputs slightly decreases the precision of the model but the recall and F1-score (see the evaluation subsection) remain high that makes the model applicable for the purposes of the project.

Aspect detection

Aspect detection component is needed to identify the main topic of an utterance to understand what people have their opinion about and what their expressed feelings referred to. Similarly to emotion recognition, the robust solutions to this task are based on deep learning approaches. However, not all of them are suitable for the project, since the results are going to be used by artists and creatives to understand their subjects from different perspectives and, therefore, aspect representations should be not only machine-readable but also humanly interpretable. For this reason, we selected an unsupervised neural attention model (He, et al. 2017) that aims at discovering diverse and coherent aspects.

The model assumes coherence by exploiting the distribution of word co-occurrences through the use of neural word embeddings that encourage words that appear in similar contexts to be located close to each other in the embedding space. In addition, an attention mechanism is used to de-emphasize irrelevant words during training, further improving the coherence of aspects.

The key elements of the model are presented in Figure 71. They include (i) an encoder for modelling the sentence embedding with an attention mechanism and (ii) two transition steps for sentence reconstruction with a special training objective and an additional regularization term. Let us consider these elements in detail.

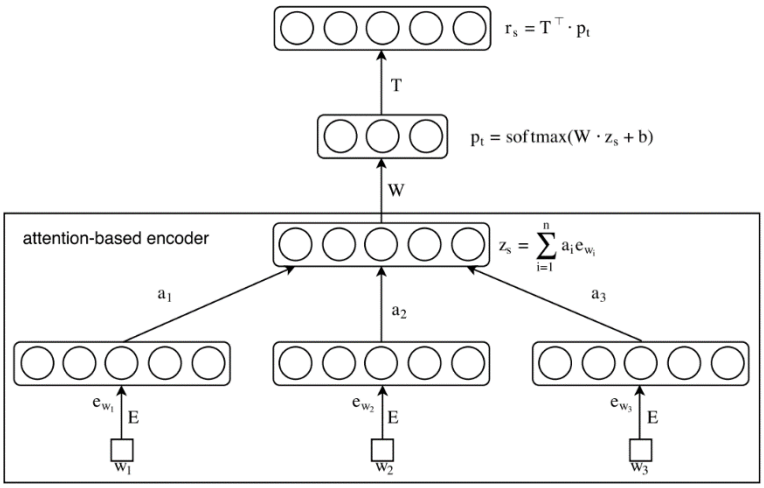


Figure 71: Architecture of neural network for aspect detection

Sentence embedding z_s is defined as the weighted summation of word embeddings e_{w_i}, where *i* corresponds to the word indexes in the sentence:

$$z_s = \sum_{i=1}^{n} a_i e_{w_i}.$$

The weight a_i is computed by an attention model, which is conditioned on the embedding of the word e_{w_i} as well as the global context of the sentence:

$$a_i = exp(d_i) / \sum_{j=1}^{n} exp(d_j),$$
$$d_i = e_{w_i}^T \cdot M \cdot y_s,$$
$$y_s = \frac{1}{n} \sum_{i=1}^{n} e_{w_i},$$

where y_s is the average of the word embeddings, which captures the global context of the sentence. $M \in R^{d \times d}$ is a matrix mapping between the global context embedding y_s and the word embedding e_w and is learned as part of the training process.

As shown in Figure 71, the reconstruction process consists of two steps of transitions, which is similar to an autoencoder. They are performed to obtain a linear combination of aspect embeddings from *T*:

$$r_s = T^T \cdot p_t,$$

where r_s is the reconstructed vector representation, p_t is the weight vector over *K* aspect embeddings, where each weight represents the probability that the input sentence belongs to the related aspect. p_t is obtained by reducing z_s from *d* dimensions to *K* dimensions and then applying a SoftMax non-linearity that yields normalized non-negative weights:

$$p_t = softmax(W \cdot z_s + b)$$

where *W*, the weighted matrix parameter, and *b*, the bias vector, are learned as part of the training process.

During the training, for each positive example, several negative examples (random sentences) are sampled from the training set. Training objective function is defined in such a way that reconstructed embedding becomes closer to the embedding of a positive example and as dissimilar as possible to negative examples. The regularization term encourages orthogonality among the rows of the aspect embedding matrix *T* and penalizes redundancy between different aspect vectors. See (He, et al. 2017) for further details.

Figure 72 provides the weights of words assigned by the attention model for some example sentences picked by (He, et al. 2017) to show that the weights learned by the model correspond very strongly with human intuition.

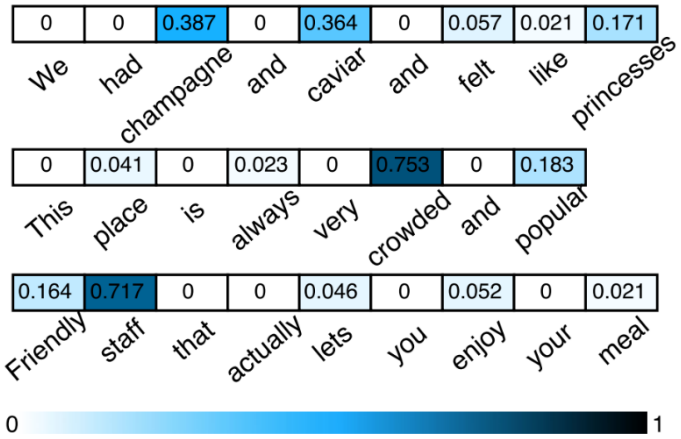


Figure 72: Visualization of the attention layer

Representative words of an aspect can be found by looking at its nearest words in the embedding space using cosine as the similarity metric. According to the authors, a long list of descriptive words (up to 50) might be considered that eases the interpretation of an aspect.

Figure 73 shows an example of grouping of representative words for the model trained using one of the collections of MindSpaces. As we can see, aspects are evenly distributed in the word embedding space without extensive overlaps that facilitates the exploration of the content of the collection.

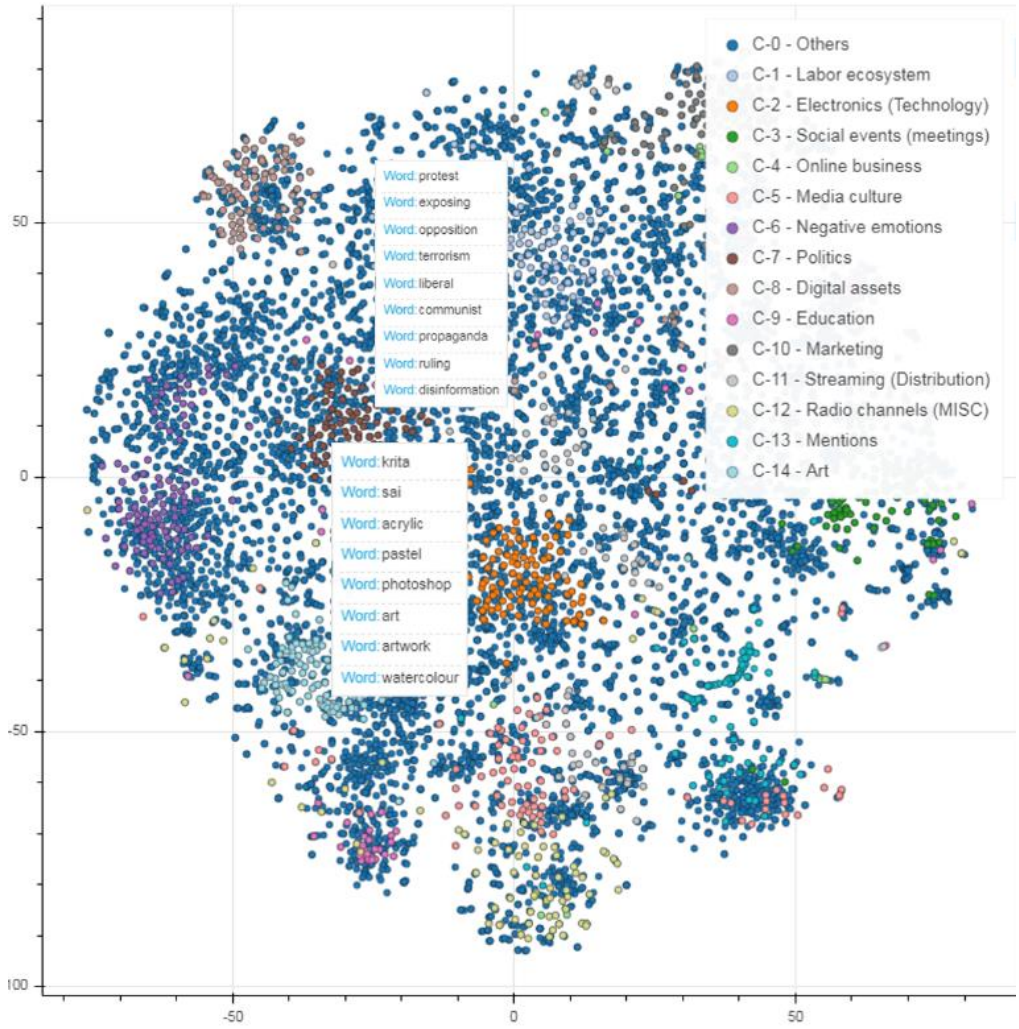


Figure 73: Representative words in a 2D projection of word-embedding space

It is worth noting that the model might be continuously trained with new coming examples and the set of aspects might evolve with time. This perfectly fits the processes in the project where textual collections are constantly updated with new materials and the possibility of periodical adjustments of the model should be envisioned.

6.6.2 Preliminary evaluation

Emotion recognition

The evaluation concerned two questions: how well the repurposed model performs on individual texts rather than on dialogues and what the distribution of emotions in the MindSpaces collections corresponding to the obtained model.

Table 20 presents the results of the evaluation on the test set from SemEval'2019 (sentiments in short dialogues). As for individual texts, the last turn in dialogue was selected for analysis as it contained the most emotive text that usually concluded and assessed the conversation. The value of the F1-score dropped insignificantly due to the decrease in precision. The recall remained to be at a high level.

Table 20: Results of the evaluation of the repurposed model

Model	Precision	Recall	F1-score
Original model for dialogues	0.65	0.82	0.73
Repurposed model for individual texts	0.61	0.83	0.70

The next step was to assess the distribution of emotions in different datasets (Table 20). The large dataset with reviews about restaurants was selected as a reference set. Collections for PUC1 and PUC3 were selected for this evaluation as they are more emotive than texts of newspapers and blogs. The copies of transcripts of interviews in French translated to English were used at this stage.

Table 21: Distribution of emotions detected in different datasets

Emotional class	Reference set	PUC1 collection	PUC3 collection
“Happy”	34465 (12%)	7993 (1.5%)	34 (3%)
“Sad”	16258 (6%)	13589 (2%)	43 (4%)
“Displeasure”	7789 (3%)	9262 (1.5%)	36 (3%)
“Others”	221373 (79%)	542413 (95%)	980 (90%)

The number of sentences with emotions is relatively smaller in MindSpaces collections than in the reference set which is enriched with many positive and negative statements about service and appearance of a place. The possible way to extract more feelings is to consider clauses instead of sentences as they are smaller, and the sentiment encapsulated in them might differ from the sentiment of the overall sentence.

Examples of the emotional tweets for the PUC1 with the classification confidence are shown in Table 22.

Table 22: Emotional sentences detected in the PUC1 collection

Emotion	Confidence	Text
“Happy”	0.75	*“so I've been doing some digital art and wow i actually feel pretty good about them”*
“Happy”	0.5	*“Great session: Becoming an awesome digital citizenship leader with @martypark and Dr. Mike Ribble at #iste19”*
“Sad”	0.74	*“i still can't believe charli xcx screaming digital noises at the end of lucky made me start crying”*

“Sad”	0.72	*“I really don’t want digital stuff. Everything is digital 😢. Think maybe i will rent a garage and call it my art studio”*
“Displeasure”	0.72	*“Remember when digital cameras overlayed the date and time in the corner of the photo. That was so stupid”*
“Displeasure”	0.94	*“And the action should be shown live on digital media. Telling them the State will crush these bastards”*

Aspect detection

The aspect detection model was applied to the collections for PUC1 and PUC2 as they contain a sufficient number of texts to train the neural network.

The results of clustering differ depending on the overall topic of the dataset and its size (see Figure 74 and Figure 75). As for collections of texts in different languages, the obtained sets of aspects were similar per PUC, therefore the stable list of aspects was defined for each use-case as presented in Table 23.

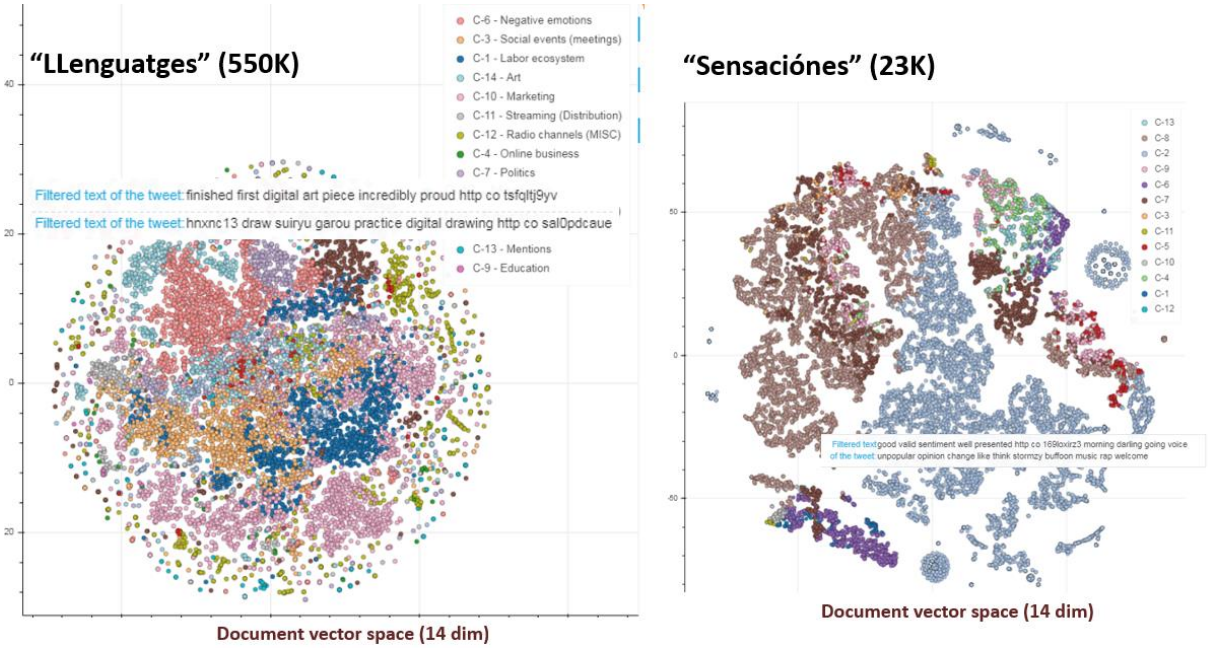


Figure 74: results of clustering for different subcollections for PUC1

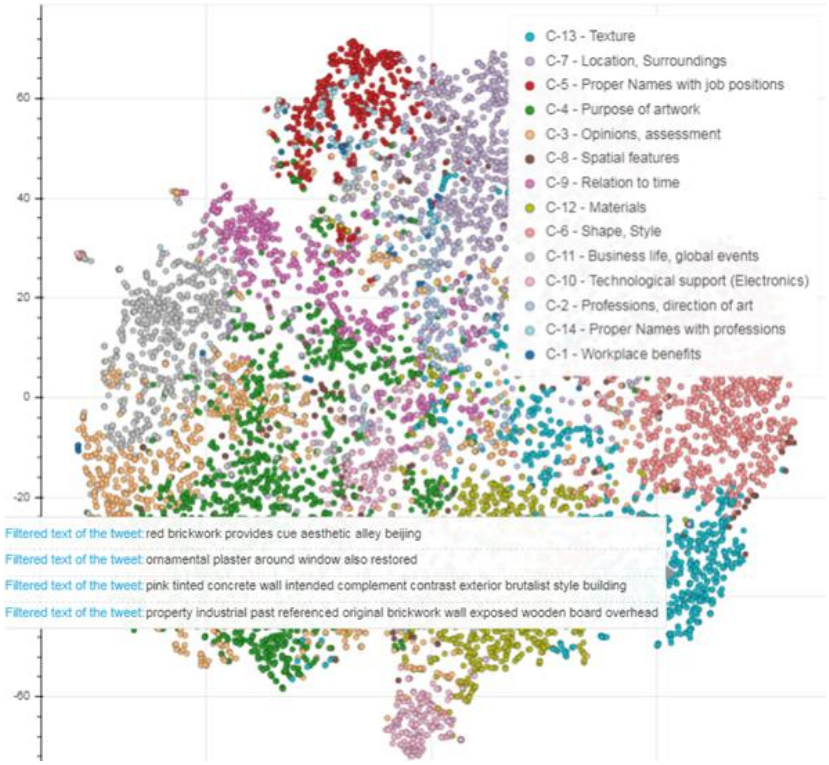


Figure 75: results of clustering for the collection of PUC2

Table 23: Defined sets of aspects grouped by the relevancy to the project

	Aspects for PUC1	**Aspects for PUC2**
Aspects relevant to the purposes of the project	- Social events, meetings; - Art; - Urban infrastructure; - Social rights, minorities; - Science, technology, and ecology	- Purpose of the product/solution; - Shape, style; - Materials; - Texture; - Spatial features; - Location, surroundings; - Technological support; - Workplace benefit
Less relevant categories to identify texts that should be discarded	- Crime and justice; - Economics and business; - Politics; - Uncategorized, irrelevant	- Business life, events; - Occupation; - Uncategorized, irrelevant

Examples of automatically extracted descriptive words for some aspects in different languages (Catalan, Spanish, English) are provided in Table 24.

Table 24: Automatically extracted aspect descriptors in different languages

Aspect	Descriptor
Urban infrastructure (in Catalan)	accessibilitat transició urbana mobilitat salut energètica bàsica eficient gestió necessitat sostenible administració qualitat millora seguretat universal governança renda anàlisi ciutadana pública econòmica comunicació mental leconomia abordar planificació connectivitat xarxa ambiental eficiència consum pràctica model criteris inclusiva indústria empresa àrea metropolitana energia transparència privada prevenció genus elèctrica responsable capacitat ecològica solucions innovació intel laboral lhabitatge producció estratègia personal millorar habitatge pobresa competitivitat ordenança activa eines avaluació garantir funció estratègic infraestructures urbà distribució visió dhabitatge provisió definir sostenibilitat mesura adequada garantia referència avançar ètica transformació activitat tecnologia impacte desenvolupament permet ligència manca reflexió econòmic centrada normativa saludable reduir transversal digitalització artificial privatització
Social events, meeting (in Spanish, stemmed words)	convoc celebr ceremoni patrocin noven asist fiest premiacion juvenil agrup festival acog barrial clausur torne lxs event concejal campus edicion todxs folclor maraton ponent fundacion feri organiz baloncest jorn sed plantel promocion regidor concej guayaquil inscrib taller patrocini provincial ix enhorabuen departamental cabild invit salon desfil estudiantil natacion acompañ certam danz soLIDARi polideport convencion jaen expositor anfitrion iberoamerican felicit inaugur bicentenari viii graduacion balonman particip une iii linar henar vii patronat ibag juventud inclusion convocatori trofe entusiasm asambl multitudinari cooperativ 1er escol sumat finaliz distrit pabellon oaxaqueñ malag apertur pre campeonat inscripcion encabez expo 4to acompañan culmin emprend precandidat content
Texture (in English)	accent toned monochromatic paintwork textural monochrome muted pale glossy pastel earthy textured detailing stucco contrasting hued tone crisp tiling contrasted herringbone lime panelling finish brickwork beige teal woodwork decor terracotta lacquer ochre terrazzo hue laminate whitewashed ornate complemented grey parquet iridescent compliment matte striped travertine stonework colouring brushed rustic teak patterned shiny velvet richly accentuated texture ribbed siding subtle juxtaposed tonal complement speckled satin palette finely uniform decoration austere subtly sleek rust minimalistic pairing contrast stain stark mahogany metalwork plaster smooth flooring turquoise paired walnut inlaid sandstone chevron enamel metallic furnishing pigmented nod dusty silvery rough mortar cabinetry tinted complementing

The obtained model might be considered to some extent as self-explainable: the representative words shown in Table 24 allow for the understanding of what the model pays more attention to when it performs further classification of new coming sentences unseen during training. Moreover, as it was mentioned earlier, the aspects might be adjusted with respect to the continuous updates in textual collections of MindSpaces that helps to capture emerging social topics reflected in texts and keep the model up-to-date.

7 REFERENCES

Afifi, M. “Dynamic length colour palettes.” *Electronics Letters* , 2019: 531-533.

Akbik, A., T. Bergmann, D. Blythe, K. Rasul, S. Schweter, and R. Vollgraf. “FLAIR: An Easy-to-Use Framework for State-of-the-Art NLP.” *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics*, 2019: 54-59.

Baziotis, C., N. Pelekis, and C. Doulkeridis. “Datastories at semeval-2017 task 4: Deep lstm with attention for message-level and topic-based sentiment analysis.” *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)* (Association for Computational Linguistics), 2017: 747–754.

Birren, F., and T Cleland. “A grammar of color: a basic treatise on the color system of Albert H. Munsell.” 1969.

Bohnet, B., and J. Nivre. “A transition-based system for joint part-of-speech tagging and labeled non-projective dependency parsing.” *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning*, 2012: 1455-1465.

Chatterjee, A., K. N. Narahari, M. Joshi, and P. Agrawal. “SemEval-2019 task 3: EmoContext contextual emotion detection in text.” *Proceedings of the 13th International Workshop on Semantic Evaluation*, 2019: 39-48.

Cohen-Or, Daniel. “Color harmonization.” *ACM SIGGRAPH Papers.* 2006. 624-630.

Cordeiro, S., C. Ramisch, and A. Villavicencio. “Ufrgs&lif at semeval-2016 task 10: rule-based mwe identification and predominant-supersense tagging.” *Proceedings of SemEval-2016*, 2016: 910-917.

Daiber, J., M. Jakob, C. Hokamp, and P.N. Mendes. “Improving efficiency and accuracy in multilingual entity extraction.” *Proceedings of the 9th International Conference on Semantic Systems (I-Semantics)*, 2013.

Devlin, J., M. W. Chang, K. Lee, and K. Toutanova. “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.” *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2019: 4171-4186.

Elliot, Andrew J., and Markus A. Maier. “Color-in-context theory.” *Advances in experimental social psychology*, 2012: 61-125.

Fader, A., S. Soderland, and O. Etzioni. “Identifying relations for open information extraction.” *Proceedings of the 2011 conference on empirical methods in natural language processing*, 2011: 1535-1545.

Frege, G. “Ueber Sinn und Bedeutung.” *Zeitschrift fuer Philosophie und philosophische Kritik* (Zeitschrift fuer Philosophie und philosophische Kritik) 100 (1892): 25-50.

Gangemi, A., V. Presutti, D. Reforgiato Recupero, A. G. Nuzzolese, F. Draicchio, and M. and Mongiovì. “Semantic web machine reading with FRED.” *Semantic Web* 8, no. 6 (2017): 873-893.

Gatys, L. A., Ecker, A. S., & Bethge, M. “ Image style transfer using convolutional neural networks.” *computer vision and pattern recognition.* IEEE, 2016. 2414-2423.

Geurts, P., D. Ernst, and L. Wehenkel. “Extremely randomized trees.” *Machine learning* 63, no. 1 (2006): 3–42.

Grammatikopoulos, Lazaros, Ilias Kalisperakis, Eleni Petsa , and Christos Stentoumis. “3D city models completion by fusing LIDAR and image data.” *Proc. SPIE 9528, Videometrics, Range Imaging, and Applications XIII, 95280O.* 2015.

Graves, A., and J. Schmidhuber. “Framewise phoneme classification with bidirectional LSTM and other neural network architectures.” *Neural networks* 18, no. 5-6 (2005): 602-610.

Gu, J., Z. Lu, H. Li, and V.O.K. Li. “Incorporating copying mechanism in sequence-to-sequence learning.” *Proceedings of the ACL*, 2016: 1631-1640.

Haller, K. “Colour in interior design.” *Colour Design*, 2017: 317–348.

Halliday, M.A.K. *Halliday's Introduction to Functional Grammar.* London & New York: Routledge, 2013.

Hasibi, F., K. Balog, and S.E. Bratsberg. “Entity linking in queries: Tasks and evaluation.” *Proceedings of the International Conference on The Theory of Information Retrieval* (ACM), 2015: 171-180.

He, R., W. S. Lee, H. T. Ng, and D. Dahlmeier. “An unsupervised neural attention model for aspect extraction.” *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, 2017: 388-397.

Hess, Wolfgang, Damon Kohler, Holger Rapp, and Daniel Andor. “Real-time loop closure in 2D LIDAR SLAM.” *Proceedings - IEEE International Conference on Robotics and Automation* 2016-June (2016): 1271-1278.

Hirschmüller, Heiko. “Accurate and efficient stereo processing by semi-global matching and mutua information.” *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005* II, no. 2 (2005): 807-814.

Honnibal, Matthew, and Ines Montani. “spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing.” 2017.

Huang, X., and S. Belongie. “Arbitrary style transfer in real-time with adaptive instance normalization.” *IEEE Int. Conf. Comput. Vis., Venice, Italy.* 2017. 1501–1510.

Isola, Phillip. “Image-to-image translation with conditional adversarial networks.” *IEEE conference on computer vision and pattern recognition.* 2017.

Itten, J. *The Art of Color: The Subjective Experience and Objective Rationale of Color.* A VNR book. Wiley, 1974.

Jakob, Wenzel, and Marco Tarini. “Instant Field-Aligned Meshes.” *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia 2015)*, 2015.

Jancosek, Michal, and Tomas Pajdla. “Multi-view reconstruction preserving weakly-supported surfaces.” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2011: 3121-3128.

Johnson, J,, A. Alahi, and L. Fei-Fei. “Perceptual losses for real-time style transfer and super-resolution.” *European conference on computer vision. Springer.* 2016. 694–711.

Khabiri, E., Y. Li, P. Mazzoleni, and D. Vadgama. “Cognitive color palette creation using client message and color psychology,".” *IBM Journal of Research and Developmen*, 2019: pp. 4:1-4:10,.

Kita, Naoki, and Kazunori Miyata. “Aesthetic rating and color suggestion for color palettes.” *Computer Graphics Forum*, 2016.

Klein, G., Y. Kim, Y. Deng, V. Nguyen, J. Senellart, and A. Rush. “Opennmt: Neural machine translation toolkit.” *Proceedings of the 13th Conference of the AMTA*, 2018: 177-184.

Labrecque, I. Lauren, and Milne George R. “Exciting red and competent blue: the importance of color in marketing.” *Journal of the Academy of Marketing Science*, 2012: 711-727.

Labrecque, Lauren I., Vanessa M. Patrick, and George R. Milne. “The marketers’ prismatic palette: A review of color research and future directions.” *Psychology & Marketing*, 2013: 187-202.

Lample, G., M. Ballesteros, S. Subramanian, K. Kawakami, and C. and Dyer. “Neural Architectures for Named Entity Recognition.” *Proceedings of NAACL-HLT*, 2016: 260-270.

Leonardis, Aleš, Horst Bischof, Axel Pinz, Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. “Computer Vision – ECCV 2006 SURF: Speeded Up Robust Features.” *Computer Vision – ECCV 2006* 3951, no. July 2006 (2006): 404-417-417.

Liu, H., P. N. Michelini, and D Zhu. “Artsy-gan: A style transfer system with improved quality, diversity and performance.” *24th International Conference on Pattern Recognition (ICPR). IEEE,.* 2018. 79–84.

Low, David G. “Distinctive image features from scale-invariant keypoints.” *International Journal of Computer Vision*, 2004: 91-110.

Luong, T., H. Pham, and C.D. Manning. “Effective approaches to attention-based neural machine translation.” *Proceedings of the EMNLP*, 2015: 1412-1421.

MacQueen, J. B. “Some methods for classification and analysis of multivariate observations.” *Fifth Symposium on Math, Statistics, and Probability.* 1967. 281–297.

Manning, Christopher D., Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard, and David McClosky. "The Stanford CoreNLP Natural Language Processing Toolkit." *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 2014: 55-60.

Martínez Alonso, H., and D. Zeman. "Universal Dependencies for the Ancora Treebanks." *Procesamiento del Lenguaje Natural* 57 (2016): 91-98.

Mel'čuk, Igor. *Dependency syntax.* Albany, NY: State University of New York Press, 1988.

Moon, P., and E. Spencer. "Geometric formulation of classical color harmony." *J. Opt. Soc. Am.*, 1944: 46–59.

Moulon, Pierre, Pascal Monasse, and Renaud Marlet. "Adaptive Structure from Motion with a contrario model estimation 1 Introduction 2 Structure from Motion the classical pipeline." *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* 7727 LNCS, no. PART 4 (2012): 1-14.

Mur-Artal, Raul, J. M.M. Montiel, and Juan D. Tardos. "ORB-SLAM: A Versatile and Accurate Monocular SLAM System." *IEEE Transactions on Robotics* 31, no. 5 (2015): 1147-1163.

Nistér, David. "An efficient solution to the five-point relative pose problem." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26, no. 6 (2004): 756-770.

Nivre, J., et al. "Universal dependencies v1: A multilingual treebank collection." *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, 2016: 1659-1666.

O'Donovan, P., A. Agarwala, and A Hertzmann. "Color compatibility from large datasets." *ACM Trans. Graph*, 2011: 1–63.

OU, L.-C., and M. R. LUO. "A colour harmony model for twocolour combinations." 2006: 191–204.

OU, L.-C., et al. "A cross-cultural comparison of colour emotion for two-colour combinations." *Color Research & Application*, 2012: 23–43.

OU, L.-C., P. CHONG, M. R LUO, and C MINCHEW. "Additivity of colour harmony." *Color Research & Application* , 2011: 355–372.

Piccinno, F., and P. Ferragina. "From TagME to WAT: a new entity annotator." *Proceedings of the first international workshop on Entity recognition & disambiguation*, 2014: 55-62.

Prokopidis, Prokopis, and Harris Papageorgiou. "Universal Dependencies for Greek." *Proceedings of the NoDaLiDa 2017 Workshop on Universal Dependencies (UDW 2017)*, 2017: 102-106.

Robin, Lennon, and Karen Plunkett-Powell. "Home Design from the Inside Out: Feng Shui, . Penguin Arkana." *Color Therapy, and Self-awareness*, 1997.

Saif, M., Mohammad, and Svetlana Kiritchenko. “WikiArt Emotions: An Annotated Dataset of Emotions Evoked by Art.” *Language Resources and Evaluation Conference.* Miyazaki, Japan, 2018.

Sanakoyeu, A., D Kotovenko, S. Lang, and B. Ommer. “A style-aware content loss for real-time hd style transfer.” *European Conference on Computer Vision.* 2018. 698–714.

Schenkel, R., F. Suchanek, and G. Kasneci. “YAWN: A semantically annotated Wikipedia XML corpus.” *Datenbanksysteme in Business, Technologie und Web (BTW 2007)–12 Fachtagung des GI-Fachbereichs, Datenbanken und Informationssysteme (DBIS)*, 2007.

Schmitz, M., S. Soderland, R. Bart, and O. Etzioni. “Open language learning for information extraction.” *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 2012: 523-534.

See, A., P.J. Liu, and C.D. Manning. “Get to the point: Summarization with pointer-generator networks.” *Proceedings of the ACL*, 2017: 1073-1083.

Shang, J., J. Liu, M. Jiang, X. Ren, C. R. Voss, and J. Han. “Automated phrase mining from massive text corpora.” *IEEE Transactions on Knowledge and Data Engineering* 30, no. 10 (2018): 1825-1837.

Sheng, L., Z. Lin, J. Shao, and X. Wang. “Avatar-net: Multi-scale zeroshot style transfer by feature decoration.” *IEEE Conference on Computer Vision and Pattern Recognition.* 2018. 8242–8250.

Simonyan, K., and A. Zisserman. “Very Deep Convolutional Networks for Large-Scale Image Recognition.” *The 3rd International Conference on Learning Representations (ICLR2015).* n.d.

Smetanin, S. “EmoSense at SemEval-2019 Task 3: Bidirectional LSTM Network for Contextual Emotion Detection in Textual Conversations.” *Proceedings of the 13th International Workshop on Semantic Evaluation*, 2019: 210-214.

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. “Dropout: A simple way to prevent neural networks from overfitting.” *Journal of Machine Learning Research* 15, no. 1 (2014): 1929–1958.

Stahlke, Samantha N., and Zaman Loutfouz. “Chromotype: A Computer-Assisted Design Tool for Palette Generation.” *CHI Conference on Human Factors.* 2018 .

Sturm, Peter. “On focal length calibration from two views.” *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2 (2001): 145-150.

SZABÓ, F, BODROGI, P., and J SCHANDA. “ Experimental modeling of colour harmony.” *Color Research & Application*, 2010: 34–49.

Ulyanov, D., A. Vedaldi, and V. Lempitsky. “Improved texture networks: Maximizing quality and diversity in feed-forward stylization and texture synthesis.” *IEEE Conference on Computer Vision and Pattern Recognition.* 2017. 6924–6932.

Wu, F., and D. S. Weld. “Open information extraction using Wikipedia.” *Proceedings of the 48th annual meeting of the association for computational linguistics*, 2010: 118-127.

Xu, Z., M. Wilber, C. Fang, A. Hertzmann, and H. Jin. “Learning from multi-domain artistic images for arbitrary style transfer.” *arXiv preprint arXiv:1805.09987.* 2018.

Yosef, M.A., J. Hoffart, I. Bordino, M. Spaniol, and G. Weikum. “AIDA: An online tool for accurate disambiguation of named entities in text and tables.” *Proceedings of the VLDB Endowment* 4, no. 12 (2011): 1450-1453.

Zhou, Qian-Yi, Jaesik Park, and Vladlen Koltun. “Open3D: A Modern Library for 3D Data Processing.” *arXiv preprint arXiv:1801.09847.*, 2018.

Zhu, J.-Y., T. Park, P. Isola, and A. Efros. “Unpaired image-to-image translation using cycle-consistent adversarial networks.” *IEEE international conference on computer vision.* 2018. 2223–2232.

8 CONCLUSIONS

Up to the first half of the project, WP4 has successfully followed the schedule of the project and supported the design and development of the MindSpaces platform. CERTH, UPF and U2M have been in close collaboration with all technological partners, but also with the artists, and architects to come up with novel ideas and implementation for supporting a new paradigm in creating and re-designing adaptive spaces.

The 3D modelling task will continue the development of a novel platform for efficiently capturing the reality. Furthermore, the some of the processes to reconstruct 3D models from captured data will be automated and a service for the semantic annotation of the point clouds will be deployed to the 2nd version on the platform.

Regarding the style transfer, a novel framework is proposed to focus on style transfer for fast and efficient unique materials' production. There are different possible options regarding the implementation of such a framework in terms of the style type that will be transferred. Some of them include the combination of a content image and more than one styles, others the combination of a content image and a collection of several artwork created from a specific artist and some other focus on the combination of a content image with another similar non art-related image in order to obtain more real-life produced output (i.e. utilizing as content a brick material image and as style another brick content image which results to a great unique very realistic output).

The textual analysis components will be improved regarding the graph-transduction grammar and the concept extraction techniques for English, Spanish, and Catalan, and will expand to French and Greek languages with all the technologies presented in this deliverable. As already discussed, statistical modules will be trained for the pre-processing step, new graph-transduction grammar will be developed, and concept extraction model will be trained. The formal text representation will be revised and improved to elaborate on the connection with the KB.

The completion of the 1st prototype of the platform is allowing partners in WP4 to get feedback from the users and carefully design the next steps for improving automation, correcting the results, and elaborating the capabilities of the platform.