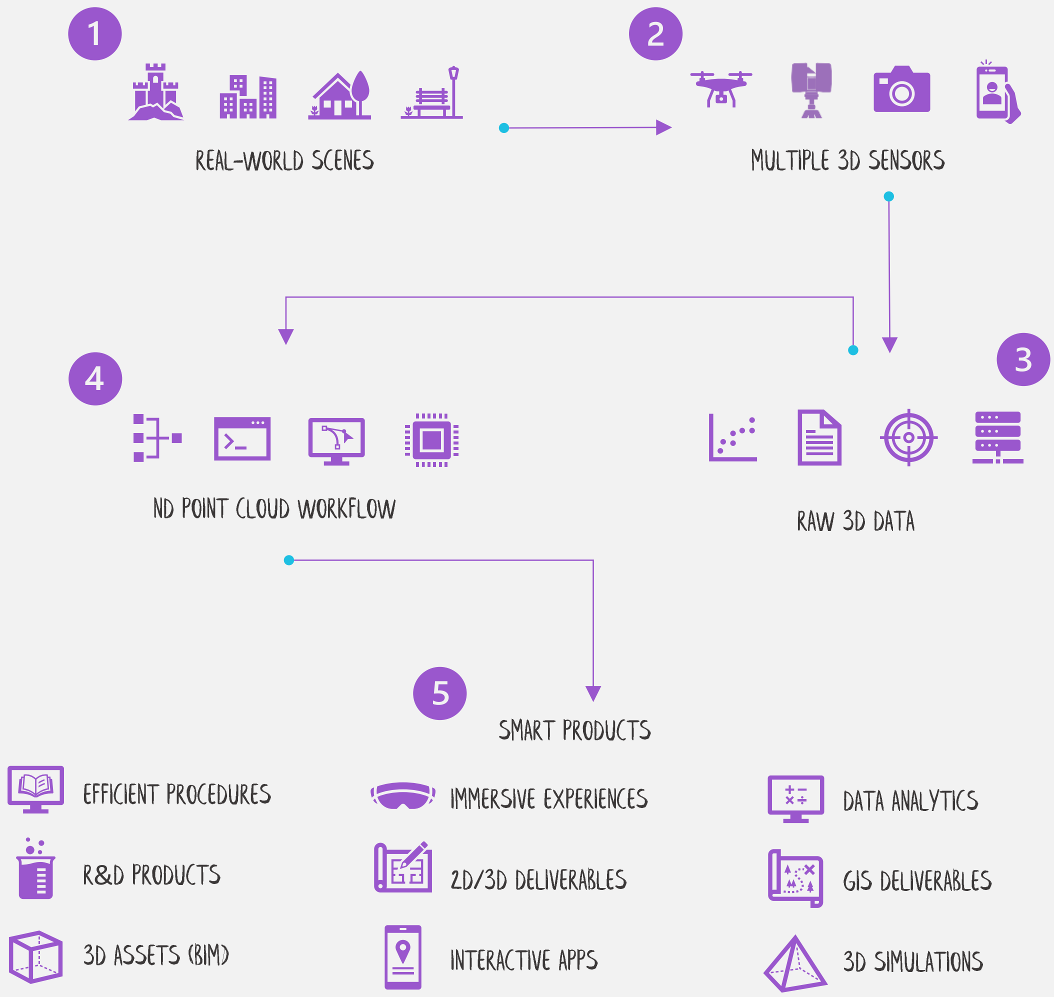
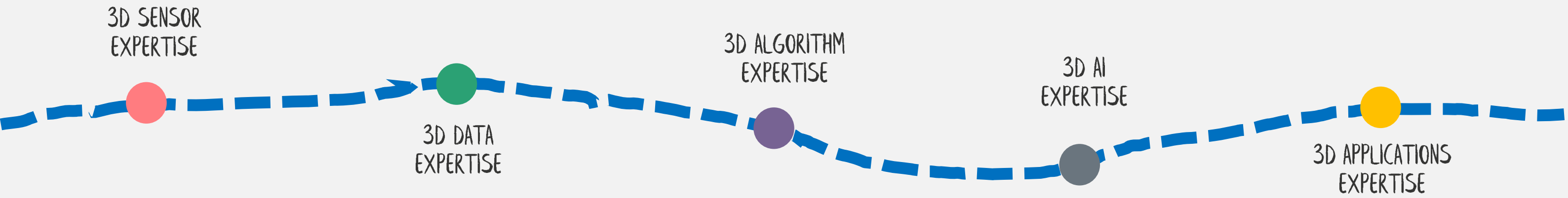
3D PERCEPTION FOR SCENE UNDERSTANDING

Discrete spatial datasets known as point clouds often lay the groundwork for decision-making applications. But can they become the next big thing? This Vision Article distills the idea generation process with 3D domain expertise: Combining Research with Development to create Solutions to Real-World challenges.





Introduction

Different renderings of a point cloud. From left to right, raw point cloud, shaded, colored, voxelized, semantized

The growing ubiquity of 3D sensors (e.g., Lidar, depth-sensing cameras, and radar) over the last few years has created a need for scene-understanding technology to process the data these devices capture. Such technology can enable machine learning (ML) and AI systems that use these sensors, like autonomous cars and robots, to navigate and operate in the real world, and can create an improved augmented reality experience on mobile devices.

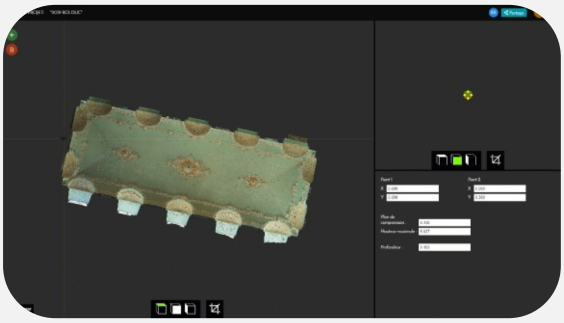
The field of computer vision has recently begun making good progress in 3D scene understanding, including models for mobile 3D object detection, transparent object detection, and more, but entry to the field can be challenging due to the limited availability tools and resources that can be applied to 3D data.

The Beginnings

I am a big point cloud enthusiast. I first discovered their existence 10 years ago, and since then, I have been tweaking my practices through the evolution of Reality Capture always to get sharper datasets. But I still remember my first surveys with terrestrial laser scanners and quickly getting these amazing (and still amazing) 3D point clouds.

I am in the process of 3D scanning an abandoned wool washing facility: 3D Viewer. © Photo R. Robroek

But then... the dream is confronted to reality. How does one effectively consider these entities? At that time, the processing—read manual overloaded repetitive digitization— was composed of several heavily manual steps such as filtering, registration, cleaning, segmenting, classifying, meshing, digitizing ... It evolved for some parts (mainly registration, filtering, and meshing) but the main bottleneck that I had back then is still unresolved: why do we bother changing the nature of the data (E.g. point cloud to vector) per application?



Manual digitization process to create a dwg file within Flyvast online point cloud software.

Is there not a more efficient workflow ?

Back almost 10 years ago, after 2 years as a 3D Laser scanning engineer, I decided to dedicate myself to teaching & research to try and solve this issue. I jumped into Academia and started investigating the current state of developments, looking for bricks that eventually need some mortar. Well, at that time, I quickly realized that no working attempt addressed the root of the problem. And that my endeavor would need more than some hours.

Let me bring you on a research journey to materialize thoughts in solutions.

The Observation

"when we open our eyes on a familiar scene, we form an immediate impression of recognizable objects, organized coherently in a spatial framework".

In 1980, Treisman defines in simple terms the complex mechanism behind our human sight-perception. For non-impaired human-being, it is often the primary source of information which our cognitive decision system can use to act on. This is extendable using our brain which quickly adapts to new surroundings and only uses the most important material captured though our eyes. In fact, the brain receives just three "images" every second, which are sorted and combined with prior knowledge to create the reality that we experience.

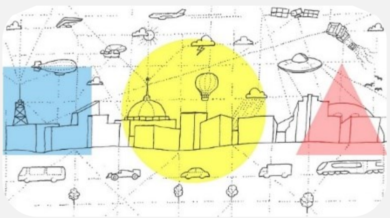
Does this image even make sense? I am sure you will find a meaning to these.

This mechanism is exceptionally fast and efficient allowing to brake when we see a red light, or simply to read this article and understand the spatial organization of words. Even more impressive, our vision can be adapted for an "orientation attention"— energy saving mode where the brain does not develop a full understanding of the surroundings—or a "discover attention"— which runs slower as the brain collects data from our memory to obtain a full understanding of the scene.

With today's computational power and high level of dematerialization, virtually replicating such a process is not only very attractive but seems feasible. While the operation is genuinely hard to mimic, studying how we interact with our environment permits to better grasp the boundaries and usable mechanisms.

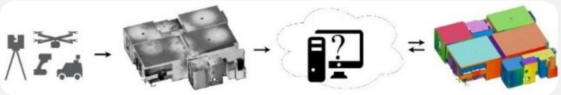
The Comparison

It first translates into using sensors that can capture key inputs usable by a computer.



Each vector in this image is guided by sensors (artificial or natural) that gather key insights for their usage.

We then aim at a procedure based on gathered data and accessible information repositories to produce a "semantic representation": a depiction of a scene integrating concepts and their meaning. In such a scenario, a spatial sensor plays the role of our eyes to obtain a digital spatial asset further refined into a semantic representation using available knowledge.



The sensor plays the role of our eyes, the spatial framework becomes a semantic representation, and the scene is tagged familiar using available knowledge.

This availability is often a first complication. Our online cognitive perception uses our memory and is structured to access needed evidence in a very short time. Mirroring this stage using a computer is extremely complex and aiming for a solution as generalist as possible is an important challenge.

The second bottleneck when trying to virtualize a cognitive decision system is the creation of a semantic representation as in the figure below. Gathering and attaching domain knowledge to underlying spatial data is linked to colossal integration and mining complications regarding data types, sources or representations.

3D point cloud representation vs 3D semantic representation

.

The Data

3D Point Clouds

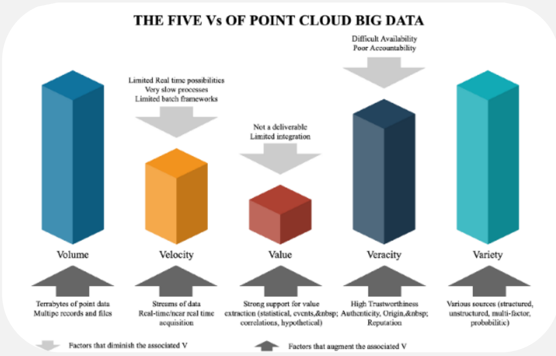
The main challenge revolves around the specificity of the data collected by the sensor(s). Single raster images or video streams are great when depth cues are not necessary but emulating our 3D visual cognition demands a richer data basis. Reality Capture devices permit to obtain such an exhaustive 3D spatial information primarily as a point cloud: a {X, Y, Z} (+ attributes) spatial ensemble which digitally represents the recorded environment w.r.t the sensor strengths and limitations.

Real-time Multi-scale point cloud of different datasets captured and combined

Point Cloud Big Data

The landscape of these instruments and acquisition methodologies is mature enough to allow digital replicas of the real world ranging from the object scale to the country scale as illustrated.

The acquisition of these so-called point clouds has become easier, faster and is even accessible from very low-cost solutions. All these hardware evolutions were unfortunately not followed by their software counterpart, which are heavily impacted by the 5 V's of Big Data problematics as illustrated below.



The Five Vs of Big Data in the context of point clouds.

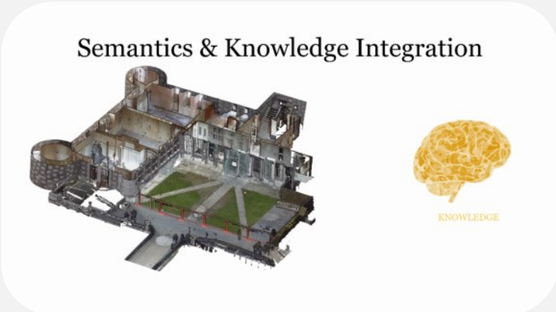
Connecting numerous sensors/approaches creates heterogeneous point cloud datasets

(Variety) and participates in the constitution of massive data repositories (Volume). In turn, it reduces the processing efficiency (Velocity) and creates new needs to turn huge amounts of point data into trustworthy (Veracity) and actionable information (Value).

THE DELIVERABLES

Point cloud acquisition and processing workflows are usually application-dependent following a classic progression from data gathering to deliverable creation. While the collection step may be specific to the sensor at hands, point-cloud-as-a-deliverable upsurges, becoming one de-facto choice for many industries. This task-oriented scenario mainly considers these as a spatial reference—which experts use to create other deliverables—thus being a project's closest link to reality. It brings accurate real-world information which could allow decision-making based on digital reality instead of interpreted or not up-to-date information.

Moreover, the procedures to convert point clouds in application-specific deliverables are very costly in time/manual intervention. It is getting ever more complicated for the human expertise to handle adequately the large and complex volumes of information, often contradictory disseminated among different actors/supports of one project.



Today, the "brain" is an expert behind a desk that will process the point cloud to extract deliverables. What we want is to integrate this knowledge directly within the data to give a semantic meaning to spatial entities

Thus, it is key for a sustainable system that big point cloud data translates into more efficient processes opening a new generation of services that help decision-making and information extraction.
We need to find ways for massive automation and structuration to avoid task-specific manual processing and non-sustainable collaboration.

THE COLLABORATION

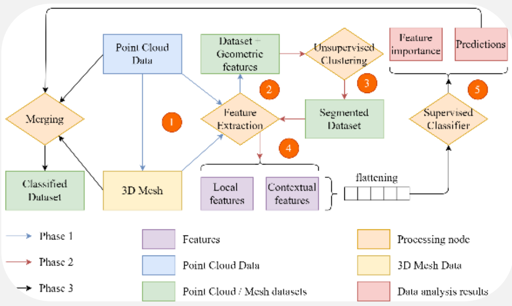
As humans, we thrive on massive collaboration. Our greatest achievements are often building on a efficient exchange of information, services and more. Point clouds are often very large depending on how much data is collected—usually in the realms of Gigabytes, if not Terabytes—and are usually destined to be archived as a reusable support to create new type of data and products. This can lead to a dead-end with exponential storage needs, incompatibility between outputs, loss of information and complicated collaboration.


This normally symbolize how we collaborate. Four fists together 😀

These practices also show limited to no attempt to generalize a framework which could in turn play as a common ground for further interoperability and generalization. This lack is counterproductive and could lead in term to a chaotic data repartition among actors and worsen the dependency to several outsourced service each aiming an application independently. This emphasize a strong need to study interoperable scenarios in which one point cloud could be used by many users from different domains, each having a different need. This will in turn introduce new constraints at the acquisition level to define the needed exhaustivity of the 3D representation for use with reasoning engines. Of course, this serializes additional challenges for interconnecting processes and insuring a compatibility with the different sources, volumes, and other data-driven parameters.

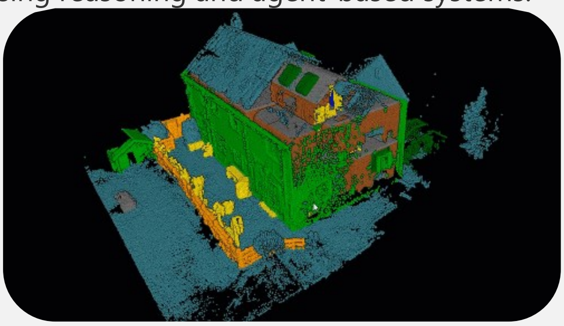
THE AUTOMATION

In this continuum, the thoughts to go from a human-centered process to an autonomous workflow orient research to develop automation and AI to speed up inference processes. This is crucial to developing point clouds in 3D capture workflows, where objects must be identified.



This is an example of a workflow proposed in a paper to combine 3D mesh with 3D point cloud predictions.

Robotics research has made a leap forward providing autonomous 3D recording systems, where we obtain a 3D point cloud of environments with no human intervention. Of course, following this idea to develop autonomous surveying means demand that the data can be used for decision-making. The collected point cloud without context does not permit to take a valid decision, and the knowledge of experts is needed to extract the necessary information and to creates a viable data support for decision-making. Automating this process for fully autonomous cognitive decision systems is very tempting but poses many challenges mainly link to Knowledge Extraction, Knowledge Integration and Knowledge Representation from point cloud. Therefore, point cloud structuration must be specifically designed to allow the computer to use it as a base for information extraction, using reasoning and agent-based systems.



Result of my AI-powered automatic 3D object recognition in an unsupervised fashion.

THE IDENTIFICATION

And this is where I want to get to: we need intelligence within our virtual datasets! This is to avoid brain work and manual processes, but also for the sake of interoperability. Many applications will use point clouds differently, but extracting deliverables per application doesn't seem to be the most efficient (and eco-friendly in terms of storage footprint). However, if experts' knowledge is formalized and integrated within point clouds, you can only guess how centralized and efficient an infrastructure becomes!

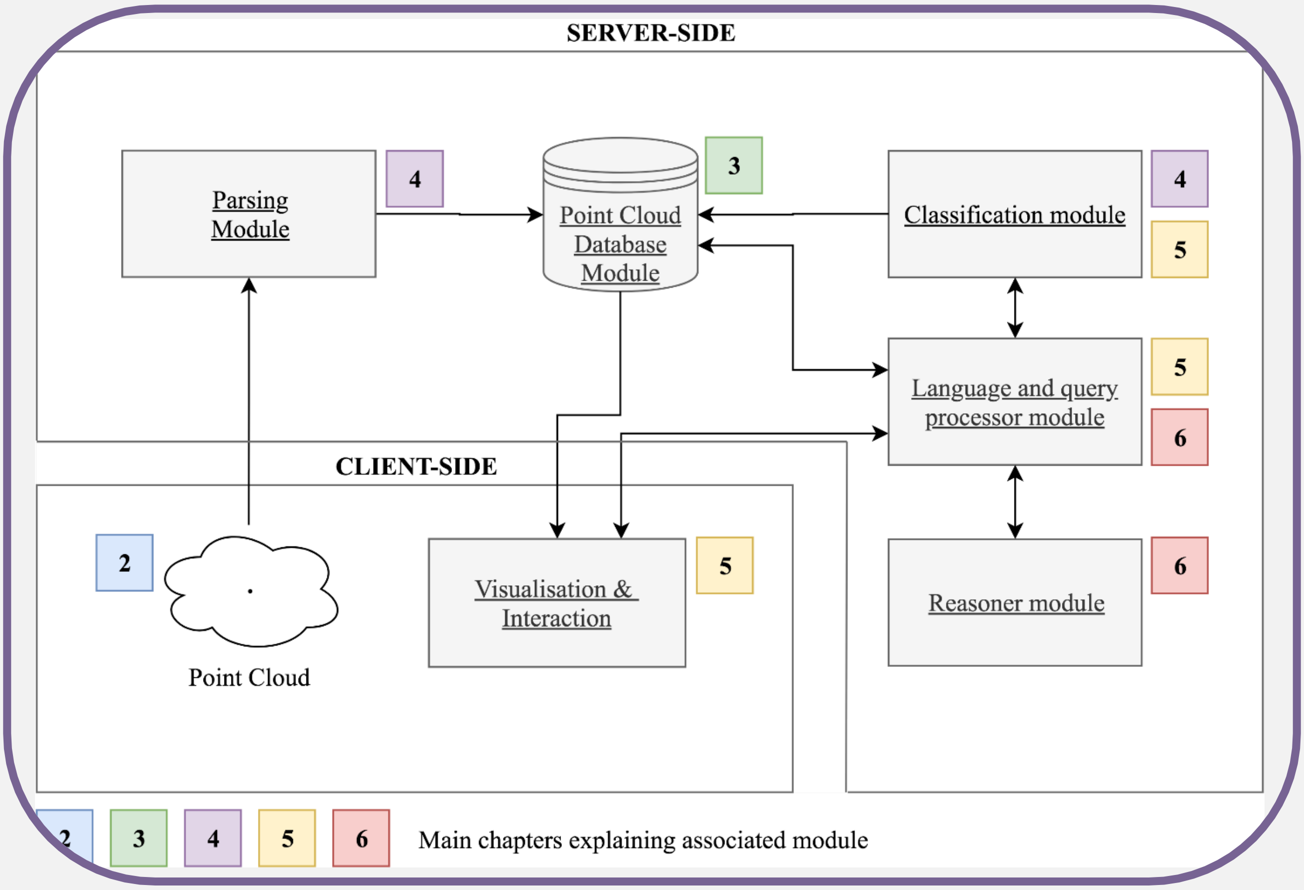
THE OUTCOME

So yes, point clouds are huge; yes, we need specific "tricks" to store them and process them, but so were videos back some decades ago! What does this imply in your specific industry? That you will soon be able to work with a "brain representation" of the 3D captured environment for you to query as you deem. But of course, the big landscape in 3D sensors makes the recognition process a wide research exploration field! Exciting!

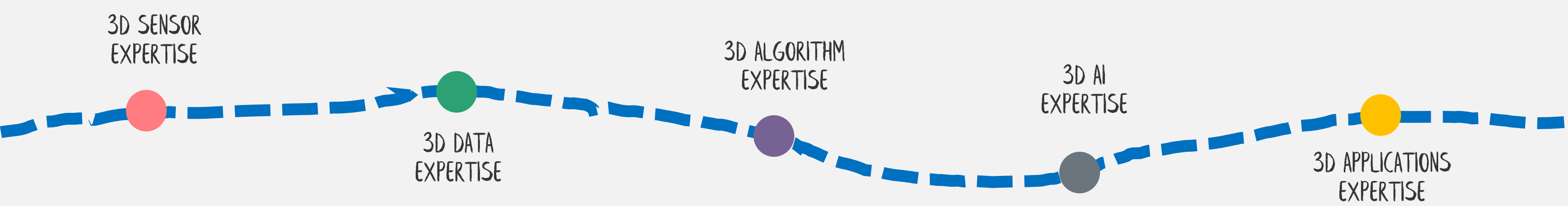
Efficient automated 3D object detection should build on these goals!

5-Point Key Takeaway

- 3D Point Clouds is close to unrefined oil coming from sensors.
- Semantic injection should aim at providing a large domain connectivity.
- The underlying data structures and algorithms are tailored to this end.
- Interoperability, modularity, and efficiency are key for collaboration



A Smart Point Cloud Infrastructure to process 3D Complex Data with Artificial Intelligence and Expert Knowledge Modelling



🎓 Learning resources

I would like to share selected learning resources for deep divers on 3D, Automation, and AI. 🤓

A. Hands-on Tutorials

I am writing open tutorials that you can find on [Medium](), targeting 3D Data Processing Workflows, combining Python and AI. At this stage, my recommended read is: [How to Represent 3D Data?]()

B. Full Standalone Courses

- 3D Python Course | [View]().
- 3D Reconstruction Course | [View]().
- 3D Point Cloud Processing Course | [👀]()
- 3D Object Detection Course | [View]().

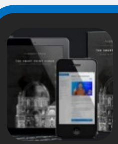
C. Research Articles (Open-Access)

- **Poux, F.**, & Billen, R. (2019). Voxel-based 3D point cloud semantic segmentation: unsupervised featuring vs. deep learning methods. [View]()
- **Poux, F.**, Ponciano, J.J., (2020). Self-learning ontology for instance segmentation of indoor point cloud. [👀]()
- **Poux, F.** et al. (2021) CityJSON building generation from airborne LiDAR 3D point clouds | [View]()
- **Poux, F.** et al. (2022). Automatic region-growing system for the segmentation of large point clouds | [View]()

Next Steps

AI-POWERED 3D PERCEPTION HAS REVOLUTIONARY POTENTIAL TO TRANSFORM INDUSTRIES, IMPROVING EFFICIENCY, ACCURACY, AND SAFETY. THIS FIRST CHAPTER OF THE 3D AI AUTOMATION COURSE IS A LIGHT INTRODUCTION WITH THE STARTING POINT: 3D POINT CLOUDS. THE COMBINATION OF AI AND 3D DATA PROCESSING CAN SAVE BUSINESSES VALUABLE TIME AND MONEY. TOMORROW, WE DIVE HEADS INTO THE BRIGHT FUTURE!

To optimize your learning journey, I engineered the Point Cloud Course: 15+ years of Research and Industry Knowledge for creating powerful 3D apps with ease! Time is running out 🎶 👉



3D POINT CLOUD COURSE

Complete course to build 3D workflows to automate 3D Point Cloud Data Acquisition, Processing, Modelling ...