Use of Generative AI for eXtended Reality

Giuseppe Amato (CNR-ISTI)
giuseppe.amato@isti.cnr.it

Extended Reality …

- Extended Reality   builds on top of

- Augmented and Mixed Reality   builds on top of

- Virtual Reality

Virtual Reality

- Users are immersed in a virtual world
- No interaction with the physical world

Augmented and Mixed Reality

- Virtual world is fused (overlayed) with physical world
- Mixed Reality users can interact with physical and virtual objects

Extended Reality

- It is generally used as an umbrella term that includes VR/AR/MR
- Sometimes it is also related to the inclusion of perception
  - Users can feel (e.g. touch) physical and virtual objects in the two fused worlds
  - Virtual objects have physical properties (material/weight/temperature…)
  - Sensors transfer data from physical to virtual world

Possible Virtual Worlds scenarios

- Imagine a group of people participating in a party
  - Some physically
  - Some using extended reality
- Local people should have the illusion that remote people are among them
- Remote people should have the illusion to be at the location of the party next to physical people

- How do we transfer the party room to the virtual world to be used by remote people?

Possible Virtual Worlds scenarios

- Immagine a group of people walking in a park
  - Some physically
  - Some using extended reality
- Remote and in presence people should have the illusion to be in the same place next to in-presence and remote people

- How do we transfer the park to the virtual world to be used by remote people?

Possible Virtual Worlds scenarios

- Imagine a doctor and a patient doing rehabilitation
  - The patient is at home
  - The doctor is in his office
  - Both make use of extended reality
- The doctor and the patient should have the illusion to be together
  - Same place
  - Possibility of physical interaction

- How do we transfer the doctor office or the patient home to the virtual world?

Possible Virtual Worlds scenarios

- Imagine a group of people attending to a conference
    - Some physically
    - Some using extended reality
- Local people should have the illusion that remote people are among them
- Remote people should have the illusion to be at the conference next to physical people

- How do we transfer the conference room, and avatar of attendees to the virtual world to be used by remote people?

Requirements and limitations

- Physical environment should be duplicated in the virtual world in a convincing way
  - Graphical appearance should be credible
  - Semantics of objects in the virtual world should be consistent with the physical one
  - Physical properties should be consistent as well

- This has a very high cost and poor scalability
  - Every new environment should be digitized and transferred to the virtual world
  - Every object should be digitized, and associated to semantic and physical properties

Generative AI for Extended Reality

- Is it possible to use Generative AI to address previous limitations?
  - Can Generative AI contribute to
    - Reducing costs for creating XR applications
    - Increase scalability and reusability of virtual worlds

  - Can Generative AI be used to
    - Go beyond graphical appearance of virtual worlds
    - Automatically associate Semantics to generated virtual objects
    - Automatically associate Physical properties to generated virtual objects

SUN: **S**ocial and h**U**man ce**N**tered **XR**

- SUN XR project

  (**S**ocial and h**U**man ce**N**tered **XR**)

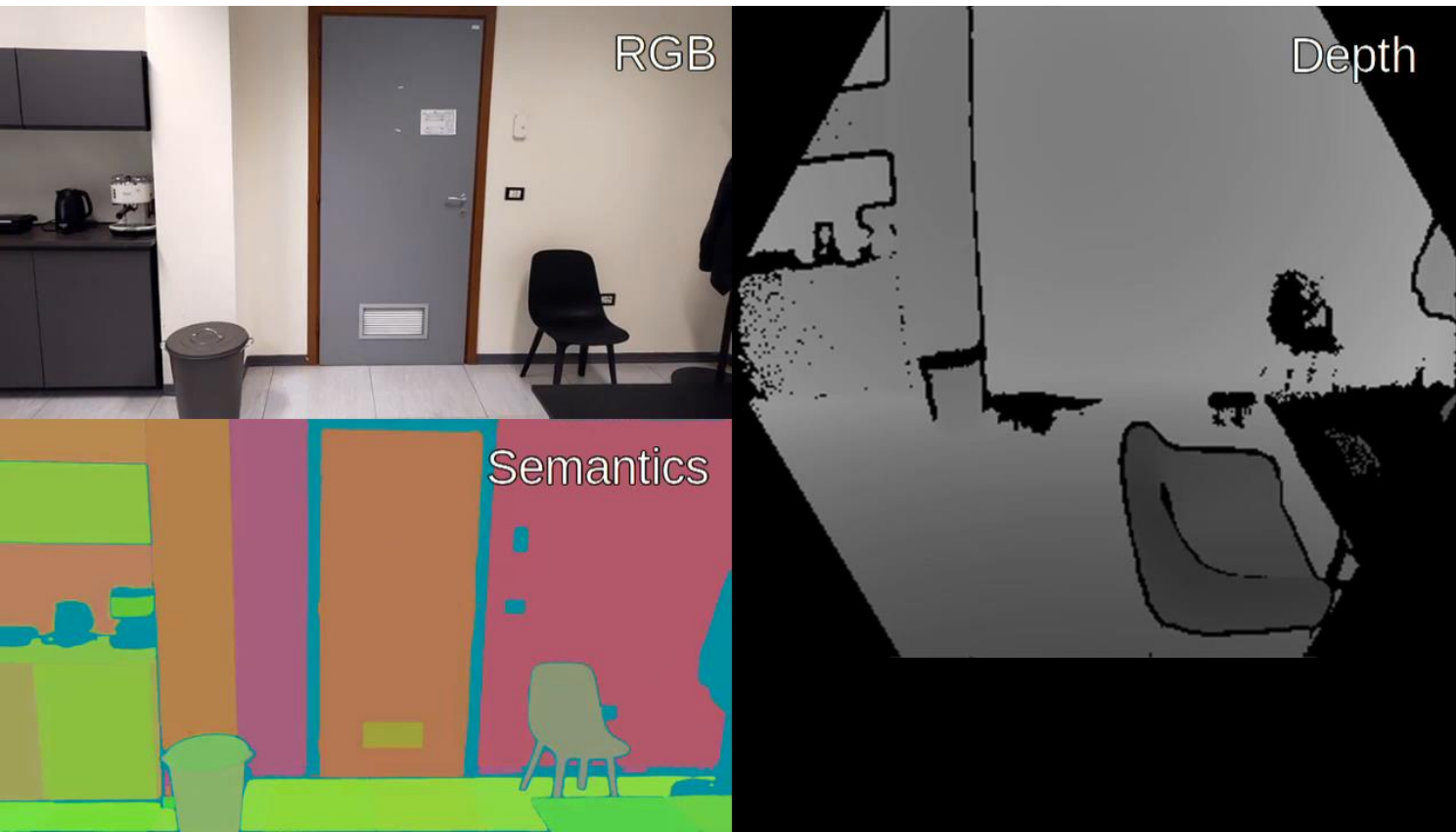 *https://www.sun-xr-project.eu/*

This project has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101092612

SUN XR Project idea in two sentences

- Address limitations in XR
  - to integrate the physical and the virtual world in a convincing way,
  - from a human and social perspective.

- The virtual world will be a means to
  - augment the physical world
  - with new opportunities for social and human interaction.

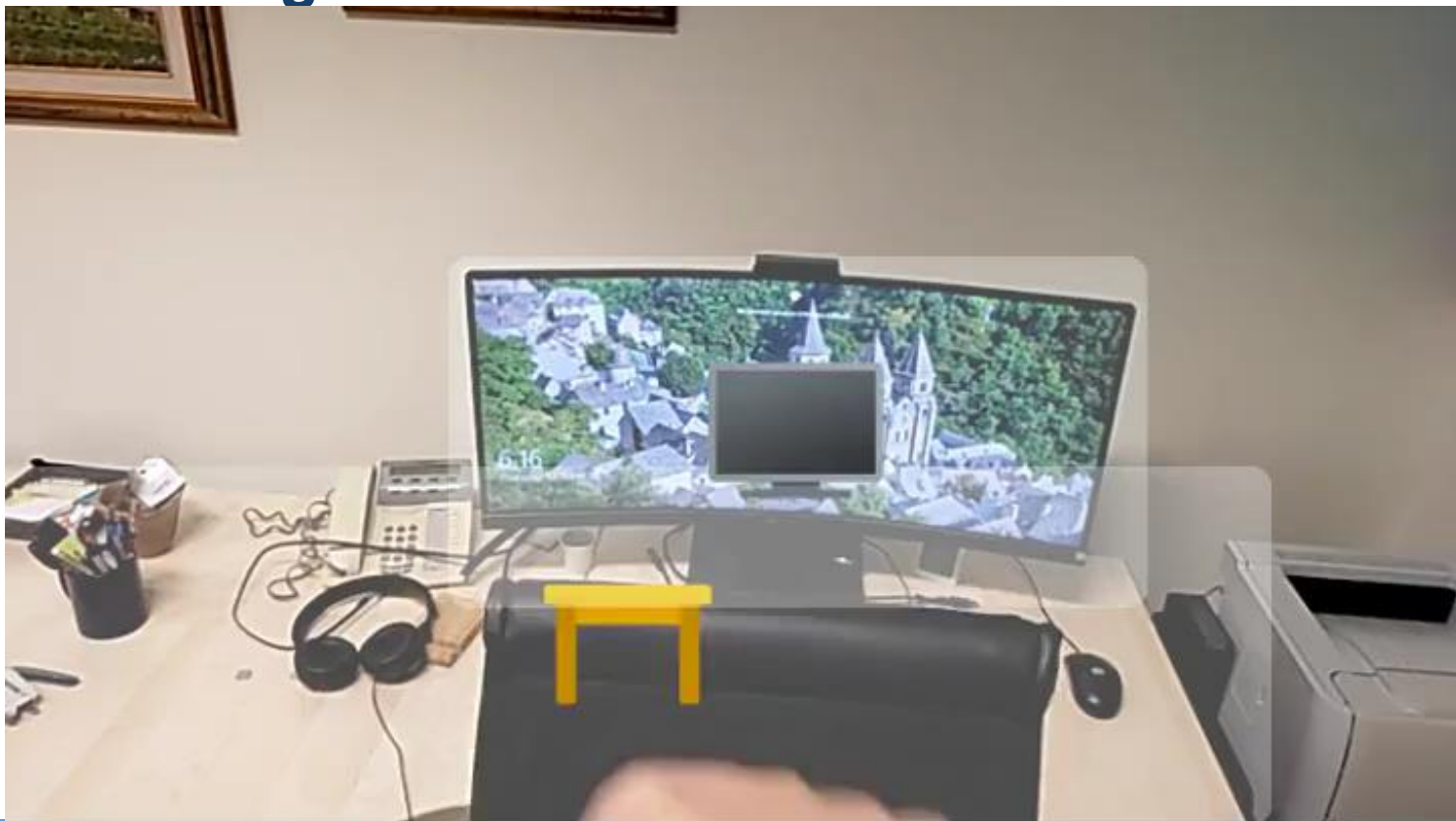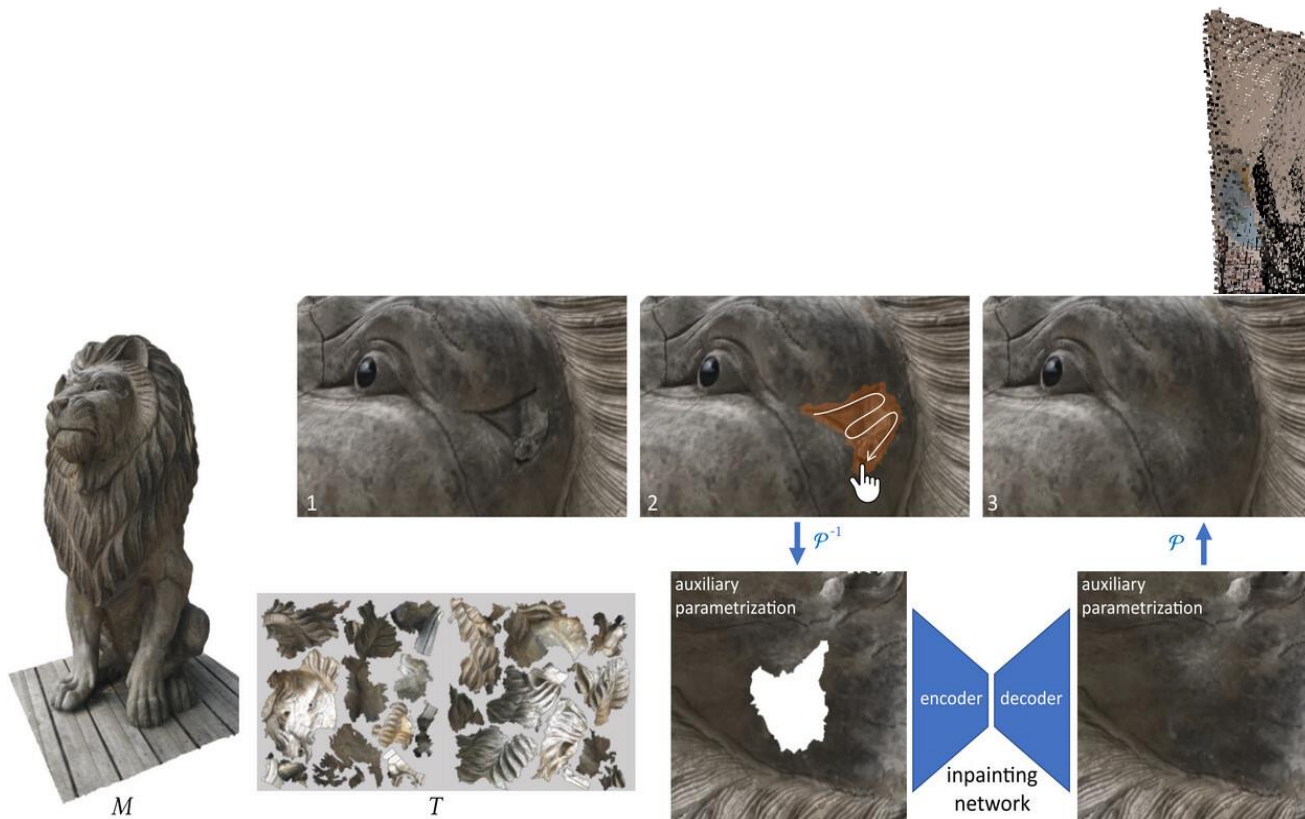# Semantic 3D Scene Reconstruction

Scene understanding

- Person 🧍
- Table 🪑
- Chair 🪑
- Etc...

Bianchi et al. "The devil is in the fine-grained details. Evaluating open-vocabulary object detectors for fine-grained understanding." CVPR 2024.

Bianchi et al. "Is CLIP the main roadblock for fine-grained open-world perception?" CBMI 2024.
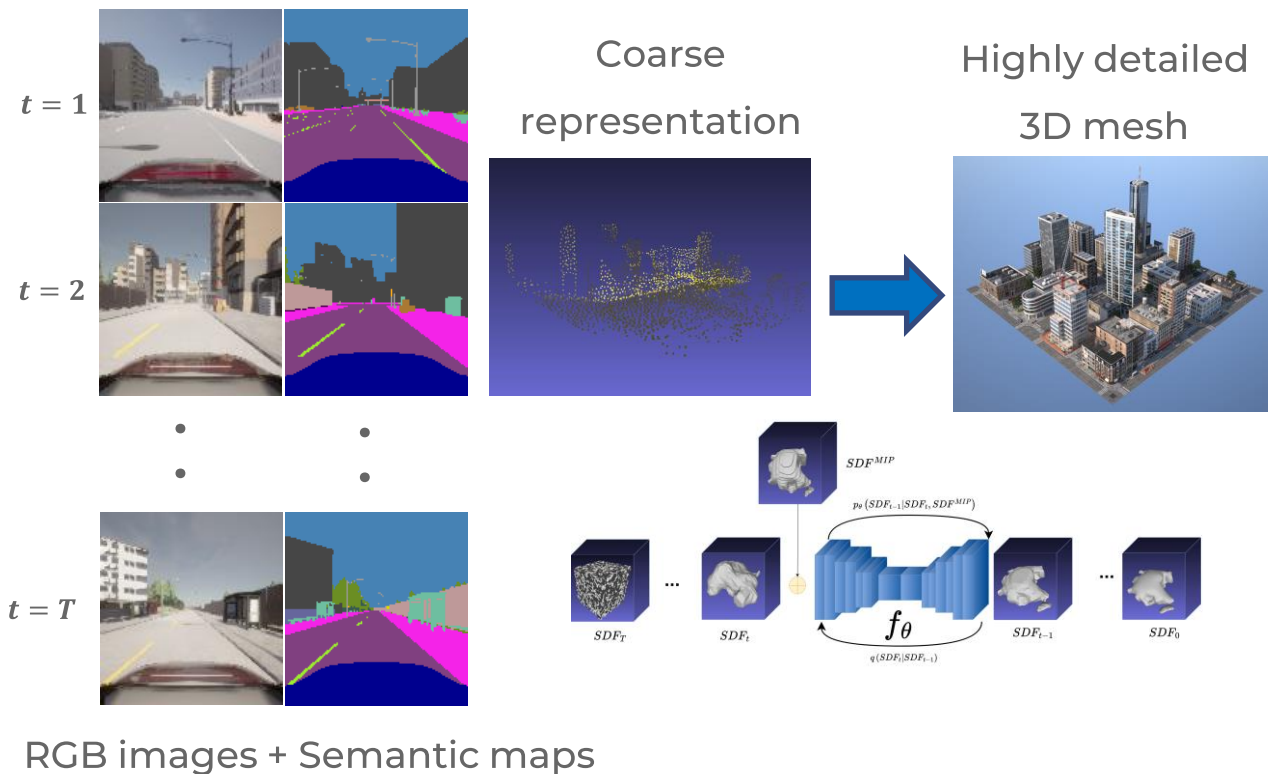
Diffusion models to correct 3D digitazion errors

Maggiordomo, A., et alt. (2023), Texture Inpainting for Photogrammetric Models. Computer Graphics Forum, 42: e14735. https://doi.org/10.1111/cgf.14735

Spatio-Temp. 3D Reconstr. using diffusion models

Spatio-Temporal 3D Reconstruction from Frame Sequences and Feature Points, Federico et alt. IMXw '24: Proceedings of the 2024 ACM International Conference on Interactive Media Experiences, https://doi.org/10.1145/3672406.36724

$t = 1$

$t = 2$

$t = T$

RGB images + Semantic maps

Coarse representation

Highly detailed 3D mesh

Promising for generating 3D models of large areas

Fully generative AI for Virtual Worlds?

Questions?

Contacts:

- *https://www.sun-xr-project.eu/*
- *Giuseppe Amato*
  - *giuseppe.amato@isti.cnr.it*