



**HAL**  
open science

# Automatic Generation of Interactive 3D Characters and Scenes for Virtual Reality from a Single-Viewpoint 360-Degree Video

Grégoire Dupont de Dinechin, Alexis Paljic

► **To cite this version:**

Grégoire Dupont de Dinechin, Alexis Paljic. Automatic Generation of Interactive 3D Characters and Scenes for Virtual Reality from a Single-Viewpoint 360-Degree Video. 26th IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR 2019), Mar 2019, Osaka, Japan. hal-02114141v1

**HAL Id: hal-02114141**

**<https://hal.science/hal-02114141v1>**

Submitted on 29 Apr 2019 (v1), last revised 27 Feb 2020 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Automatic Generation of Interactive 3D Characters and Scenes for Virtual Reality from a Single-Viewpoint 360-Degree Video

Grégoire DUPONT de DINECHIN  
Alexis PALJIC



gregoire.dupont\_de\_dinechin@mines-paristech.fr  
alexis.paljic@mines-paristech.fr

ID: 053

## Context

We want to capture a real-world scene for interactive VR. What are the pros and cons of using a 360-camera mounted on a tripod?

- |  |   |
|--|---|
| <p><b>(+) Pros</b></p> <ul style="list-style-type: none"> <li>• Easy, quick, low-cost</li> <li>• High-fidelity visual result</li> <li>• All-around vision = perfect for seated VR viewing</li> </ul> | <p><b>(-) Cons</b></p> <ul style="list-style-type: none"> <li>• No motion parallax = not adapted for moving in VR</li> <li>• Very limited potential for adding interactivity</li> </ul> |
|--|---|

The idea is to keep the pros and eliminate the cons, by automatically estimating 3D data from the input 360-degree video.

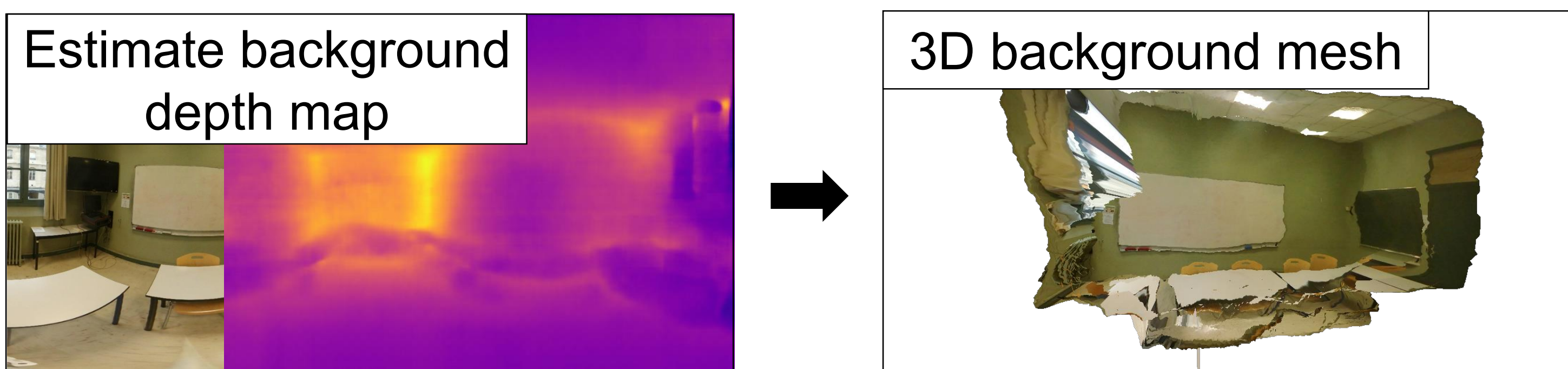
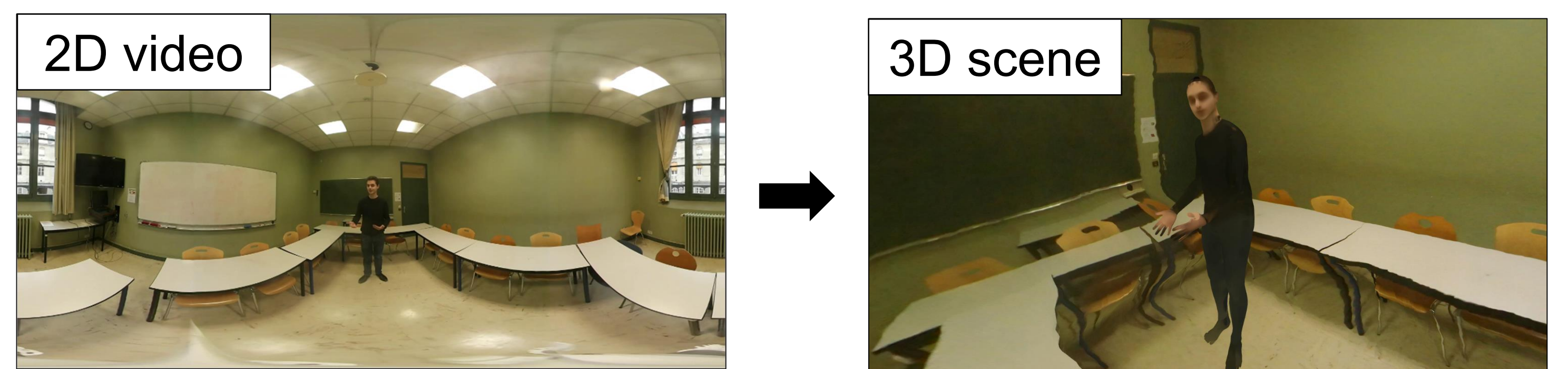
## Contributions

- A complete and automatic 3D scene generation process that leverages available pre-trained deep learning models.**  
Input: a single-viewpoint 360-degree video.  
Output: a 3D scene with animated, textured character models.
- A user study evaluating our approach.**  
Focus: Users' reaction to the generated characters.  
Results: Virtual agents were generally perceived as convincing. Quality of character textures must be improved however.

## Scene generation process

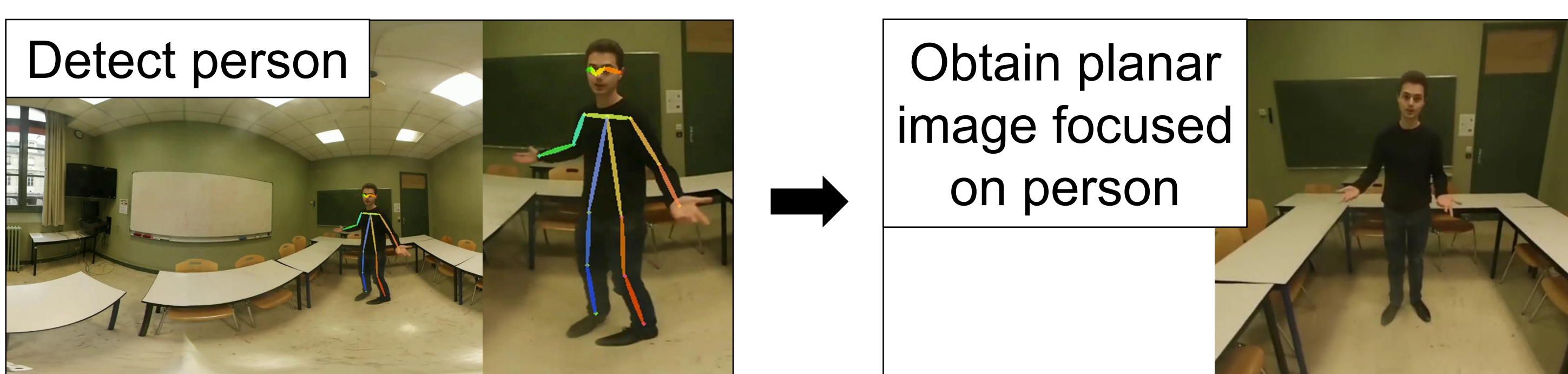
### Process overview

1. Estimate depth mesh for background.
2. Detect characters to obtain planar images for next step.
3. Estimate characters' shapes, animations and texture maps.
4. Add interactive objects and responsive character behavior.



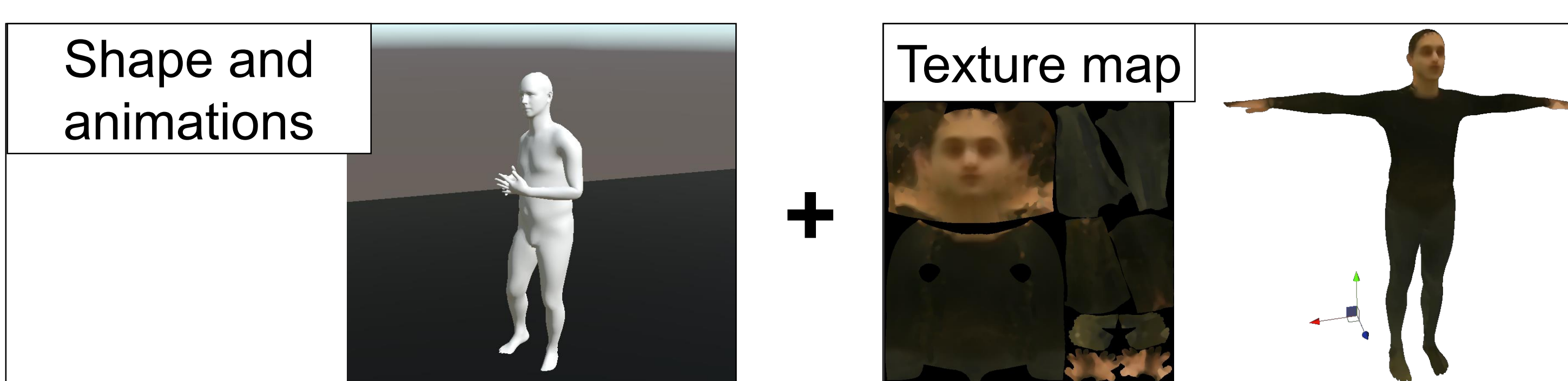
### 1. Estimate 3D background mesh

- Estimate 360-degree depth map with pre-trained model (here: *OmniDepth UResNet*).
- Create 3D triangle mesh from depth map, textured with background image.



### 2. Detect characters to obtain planar images (necessary because models in the next step do not work well on spherical images)

- Estimate characters' bounding boxes in image, with pre-trained model (here: *AlphaPose*) or background subtraction.
- Create planar images centered on characters, by perspective projection on a virtual pinhole camera.



### 3. Estimate characters' shapes, animations and texture maps

- Estimate shape and pose with pre-trained model (here: *Human Mesh Recovery*). Use poses and the virtual pinhole cameras from step 2 to obtain characters' animations in absolute 3D space.
- Estimate texture map with pre-trained model (here: *DensePose*), and optionally clean/complete it with image processing.

## User study

### Hypothesis

Users' reaction to the generated character models should be similar to their reaction to computer-generated virtual agents. Specifically, enhancing a character's behavior to make it seem more responsive should result in the character being treated more like a social object.

### Variables

- Independent: level of interactivity of the virtual agent (e.g. reaction to the user's actions).
- Dependent: behavioral measures (e.g. interpersonal distance) and self-report measures (e.g. standard questionnaire for virtual agent evaluation).

### Results (50 participants)

- Users felt that they initiated more interaction, felt more looked at, and moved closer to the agent when the character presented responsive traits.
- Users appreciated the characters' animations, but often criticized the texture map, especially the fact that it blurred important facial features.

