Ital-IA
ITALIA INTELLIGENZA ARTIFICIALE
cini National Lab AIIS

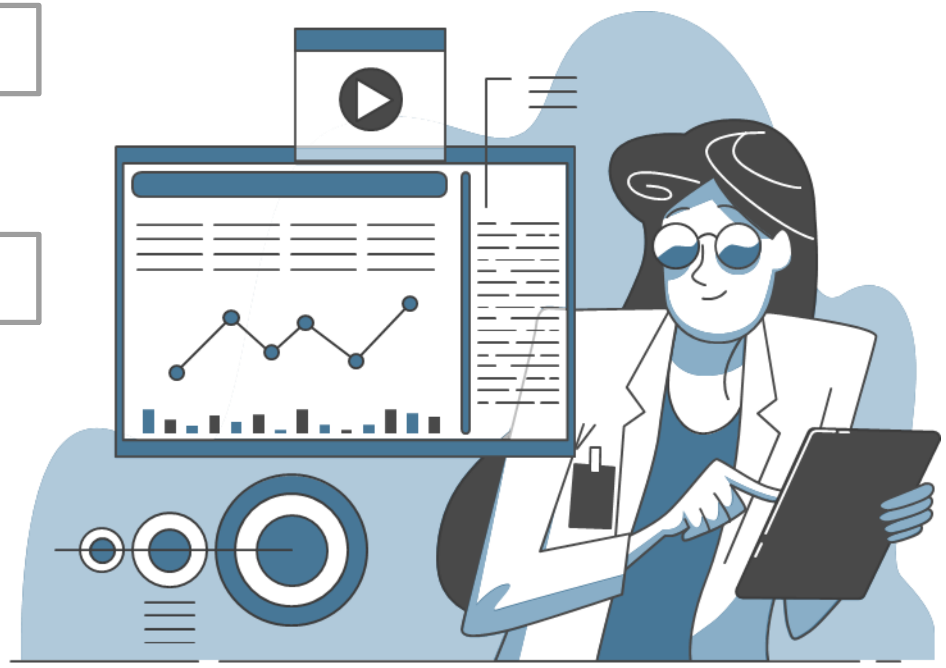
AI for Media & Games

Revolutionizing Media and Gaming with AI: Advancements in Body Measurement Calculation, Motion Tracking, Gesture Recognition, and Upper Limb Segmentation

Gilda Manfredi, Nicola Capece, Ugo Erra and Monica Gruosso

UNIVERSITÀ DEGLI STUDI DELLA BASILICATA 1982

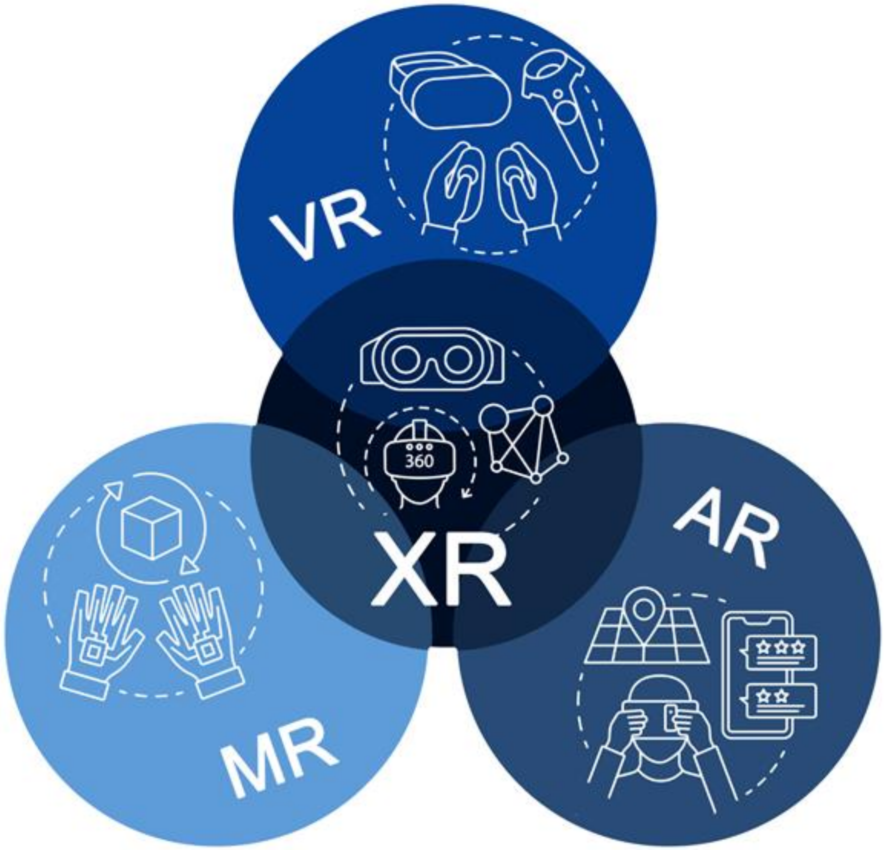
Table of Contents



"Artificial Intelligence (AI) this century is likely to be sufficiently **transformative** to precipitate a transition comparable to (or more significant than) the agricultural or industrial revolution."

Allan Dafoe, 2018

AI for Media and games



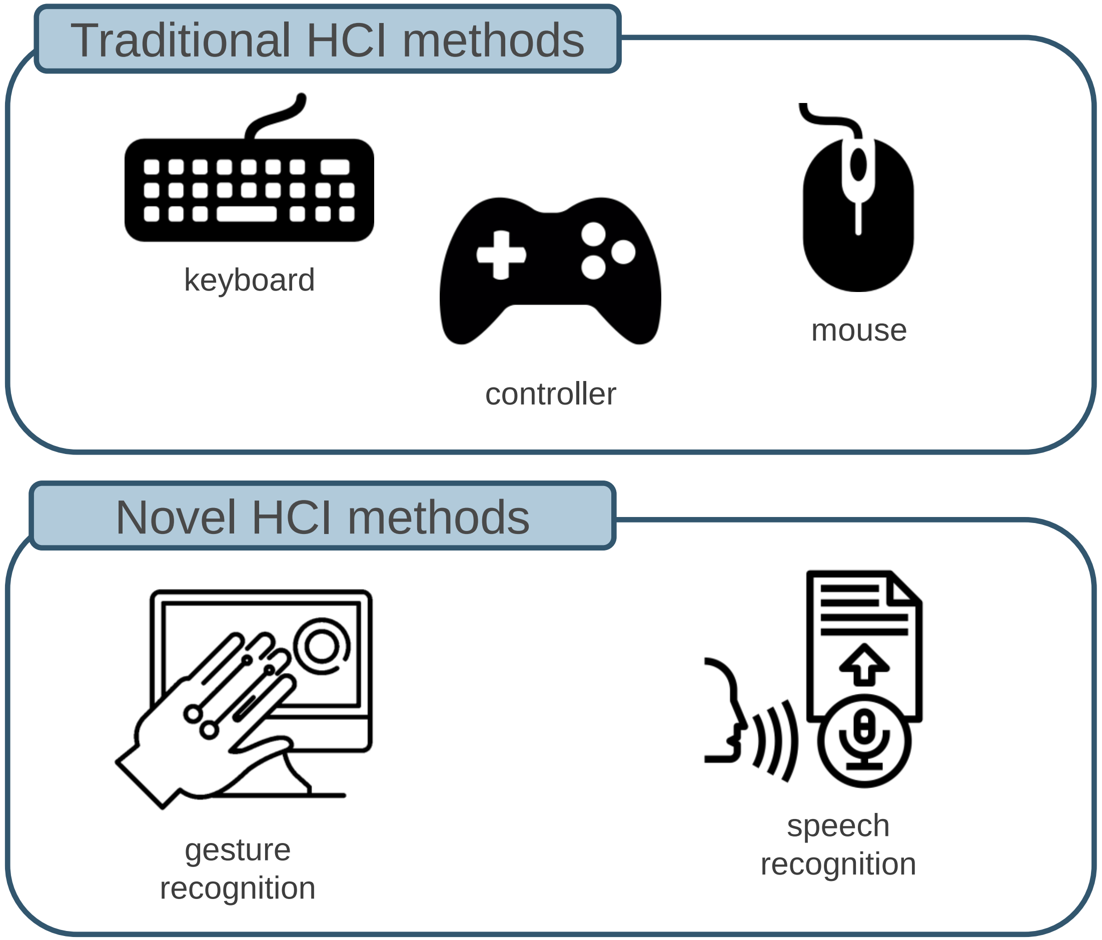
Research has shown that AI has tremendous potential in the field of media and gaming, particularly in the development of **eXtended Reality (XR)** applications.

This has led to a search for new and innovative ways to engage and immerse users.

AI for Media and games

However, to enhance the user experience in XR applications, researchers are exploring new ways of interaction that go beyond traditional Human Computer Interaction (HCI) methods.

This is due to the limitations of traditional HCI methods, which can limit the ability of users to fully immerse themselves in XR environments.



AI for Human-Computer interaction

Research activities conducted by the Laboratory of Computer Graphics and Parallel Computing (GPU Lab) of the University of Basilicata.

XR gesture recognition system

Utilizes machine learning algorithms to recognize and interpret hand gestures in real-time within virtual and augmented reality environments.

Upper limb segmentation system

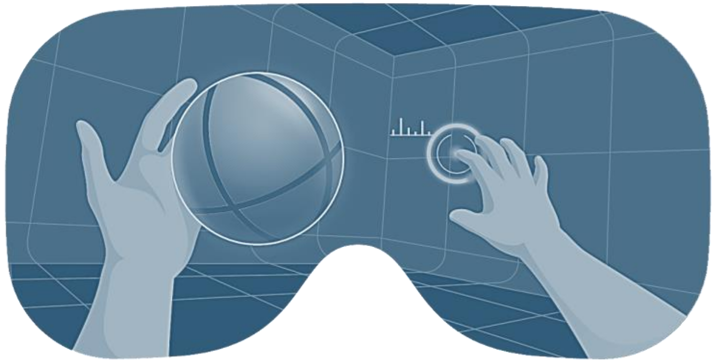
Utilizes deep learning techniques for upper limb segmentation in egocentric vision and unconstrained real-world scenarios.

Virtual Dressing Room

Enables users to create a realistic 3D model of themselves, which can be used to try on virtual clothing and accessories. It includes Body Tracking and Anthropometric Measurement Systems.

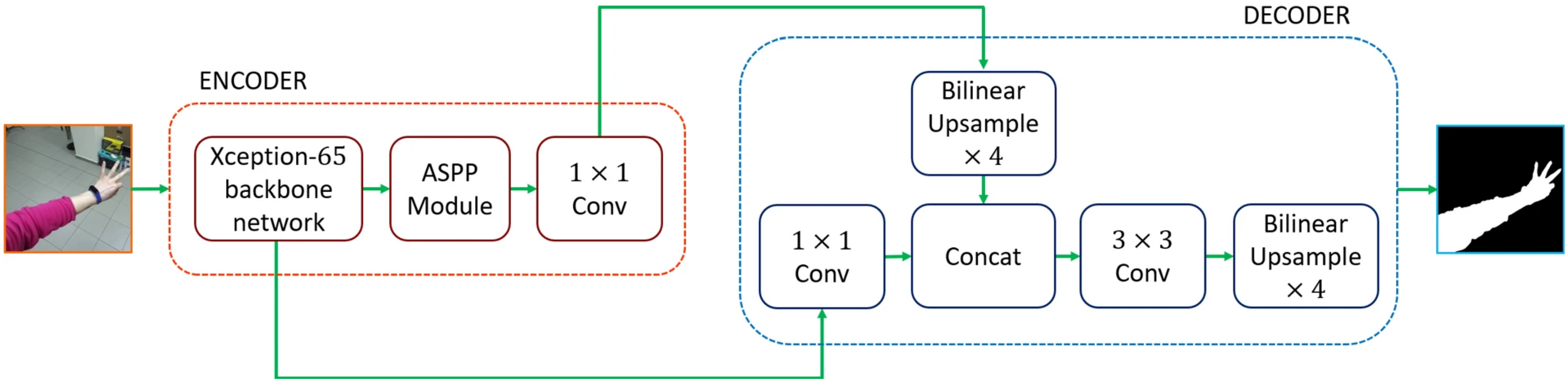
Egocentric Upper Limb Segmentation in Unconstrained Real-Life Scenarios

One promising area of research in HCI is egocentric vision-based approaches for controlling virtual avatars using body movements.



Objective	Methodology
Extend hand segmentation to upper limb segmentation in egocentric vision and real-world scenarios.	Encoder-decoder deep convolutional neural network using DeepLabv3+ architecture.

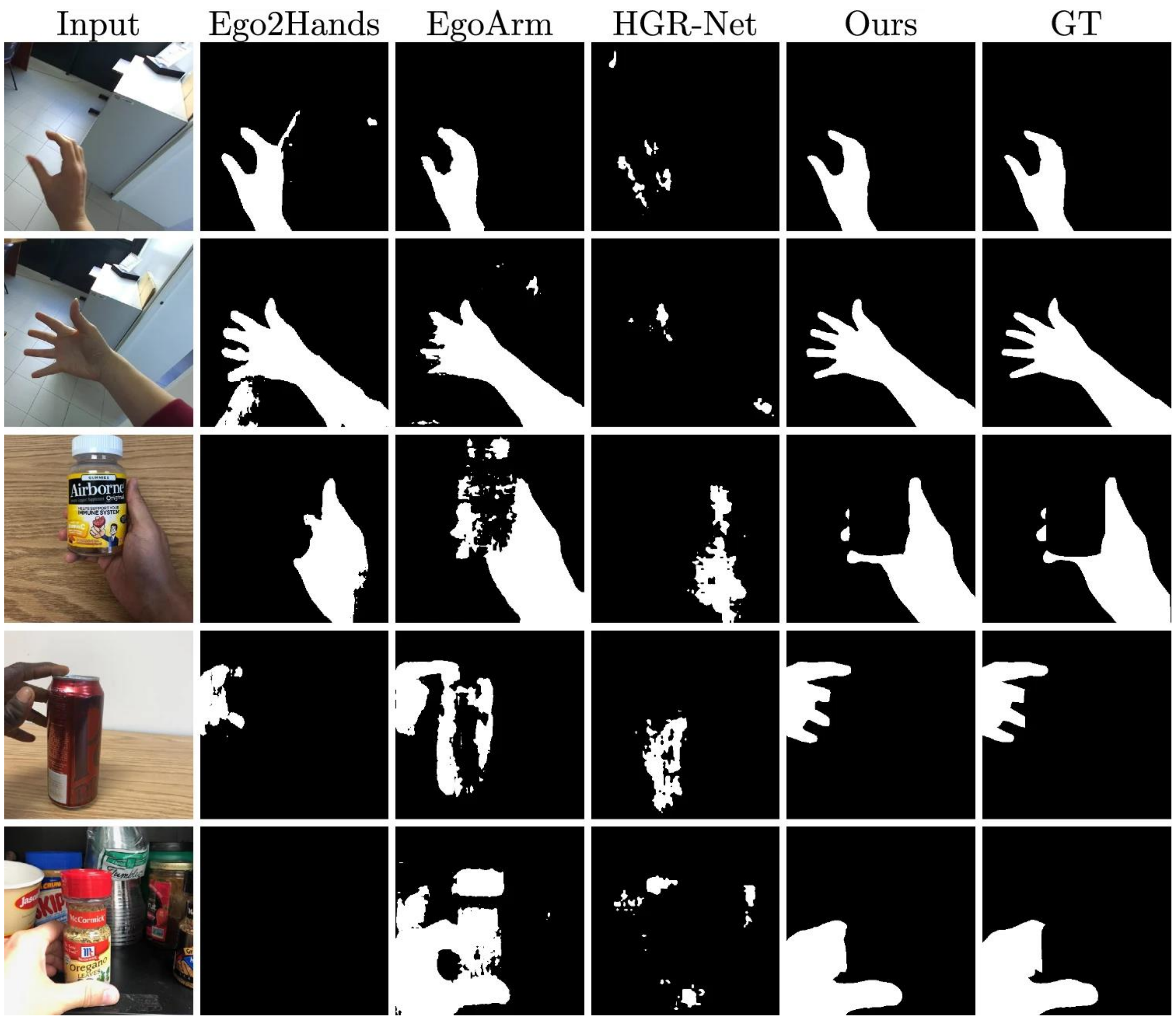
Neural Network Architecture



Dataset: Approximately 46,000 RGB images with accurate labels capturing unconstrained real-life activities.

Training: Stochastic gradient descent optimization, pre-trained weights from ImageNet and MS-COCO datasets.

Results



Our trained network achieved impressive results for both whole upper limb and hand-only segmentation tasks, significantly outperforming the state-of-the-art (SOTA).

We compared our outcomes with Ego2Hands, EgoArm, and HGR-Net, showing the superiority of our network in various challenging scenarios.

Publication

*M. Gruosso, N. Capece, U. Erra, **Egocentric upper limb segmentation in unconstrained real-life scenarios**, Virtual Reality (2022) 1–13.*

Virtual Reality
https://doi.org/10.1007/s10055-022-00725-4

S.I. : NEW TRENDS ON IMMERSIVE HEALTHCARE

Check for updates

Egocentric upper limb segmentation in unconstrained real-life scenarios

Monica Gruosso[1] · Nicola Capece[1] · Ugo Erra[1]

Received: 26 April 2022 / Accepted: 18 November 2022

Abstract
The segmentation of bare and clothed upper limbs in unconstrained real-life environments has been less explored. It is a challenging task that we tackled by training a deep neural network based on the DeepLabv3+ architecture. We collected about 46 thousand real-life and carefully labeled RGB egocentric images with a great variety of skin tones, clothes, occlusions, and lighting conditions. We then widely evaluated the proposed approach and compared it with state-of-the-art methods for hand and arm segmentation, e.g., Ego2Hands, EgoArm, and HGRNet. We used our test set and a subset of the EgoGesture dataset (EgoGestureSeg) to assess the model generalization level on challenging scenarios. Moreover, we tested our network on hand-only segmentation since it is a closely related task. We made a quantitative analysis through standard metrics for image segmentation and a qualitative evaluation by visually comparing the obtained predictions. Our approach outperforms all comparing models in both tasks and proving the robustness of the proposed approach to hand-to-hand and hand-to-object occlusions, dynamic user/camera movements, different lighting conditions, skin colors, clothes, and limb/hand poses.

Keywords Semantic segmentation · Image processing · Subtraction techniques · Neural networks · Egocentric vision

1 Introduction

Hands are one of the main channels of human communication that allows people to relate to each other and interact with objects. Our hands are often in our field of vision, for example, during daily activities (Pirsiavash and Ramanan 2012). In many cultures, hands support verbal communication and increase comprehensibility by adding meaning and emphasis to words (Maricchiolo et al. 2005). Many applications involving the use of the hands are based on localization methods and, in particular, on hand segmentation. It is usually used as a pre-processing step in various contexts, such as hand gesture recognition (Poularakis and Katsavounidis 2015), human–robot interaction (HRI) (Ju et al. 2017), human–computer interaction (HCI) (Maurya et al. 2018), and mixed reality (MR) (Herumurti et al. 2017).

With the spread of wearable devices, systems for analyzing and detecting hands in a first-person perspective, called egocentric or first-person vision (FPV), have increasingly developed (Betancourt et al. 2015). Although several approaches were designed to tackle this task using the third-person point of view (TPV) (Matilainen et al. 2016; Bojja et al. 2019), a significant difference in the visual aspect compared to the egocentric view can be found. The FPV combines the challenges of segmentation (Kok and Chan 2016; Ren et al. 2020; Gruosso et al. 2021a; Minaee et al. 2021), mainly due to a large variety of backgrounds, presence of shadows and occlusions, with the inherent difficulties in ego vision (Alletto et al. 2015), which involves rapid changes in the lighting conditions of the captured scene, dynamic movement of the camera and the wearer that can cause motion blur (Li and Kitani 2013).

In this study, we extended hand segmentation task focusing on upper limb in egocentric vision and unconstrained real-world scenarios, where not only the hand but

Monica Gruosso, Nicola Capece and Ugo Erra have contributed equally to this work.

✉ Nicola Capece
nicola.capece@unibas.it

Monica Gruosso
monica.gruosso@unibas.it

Ugo Erra
ugo.erra@unibas.it

1 Department of Mathematics, Computer Science and Economics, University of Basilicata, Via dell'Ateneo Lucano 10, Potenza 85100, Italy

Published online: 03 December 2022

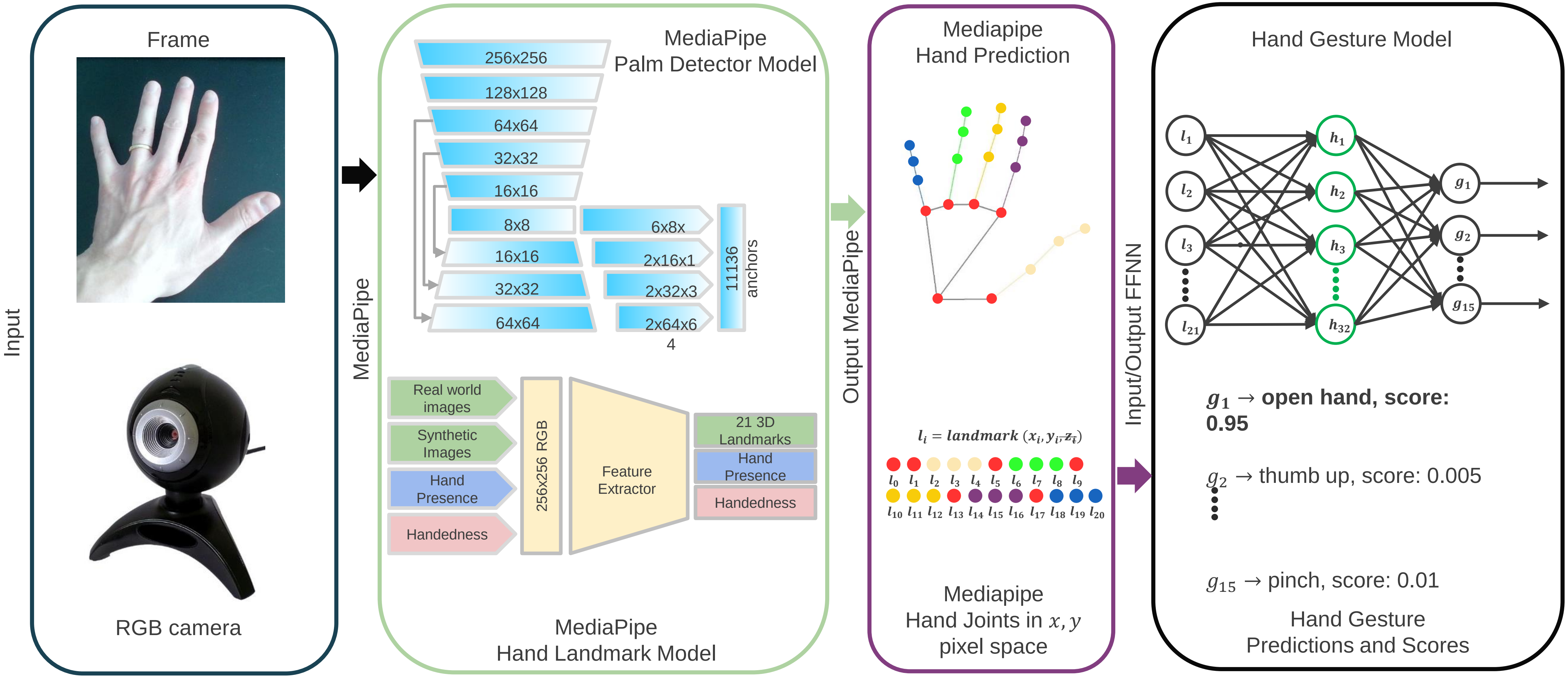
Springer

XR Hand Gesture Recognition System with a simple RGB camera

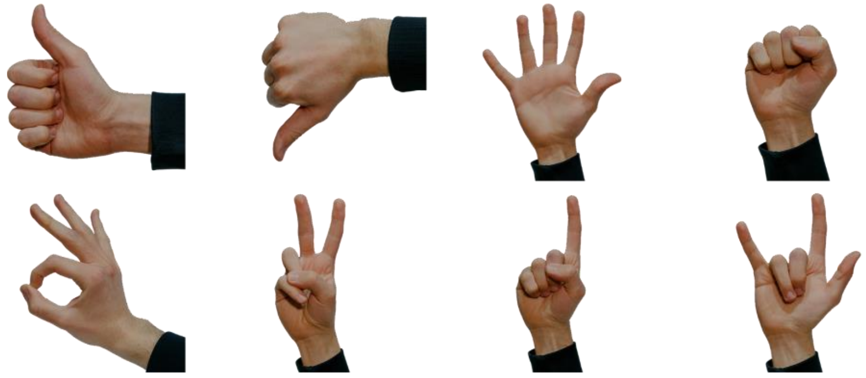
Idea: To develop a gesture tracking system that is more affordable and usable compared to Leap Motion or modern headsets.

To address the need for low-cost HGR solutions, we propose a deep learning approach that enables hand gesture recognition using a simple, low-cost RGB camera.

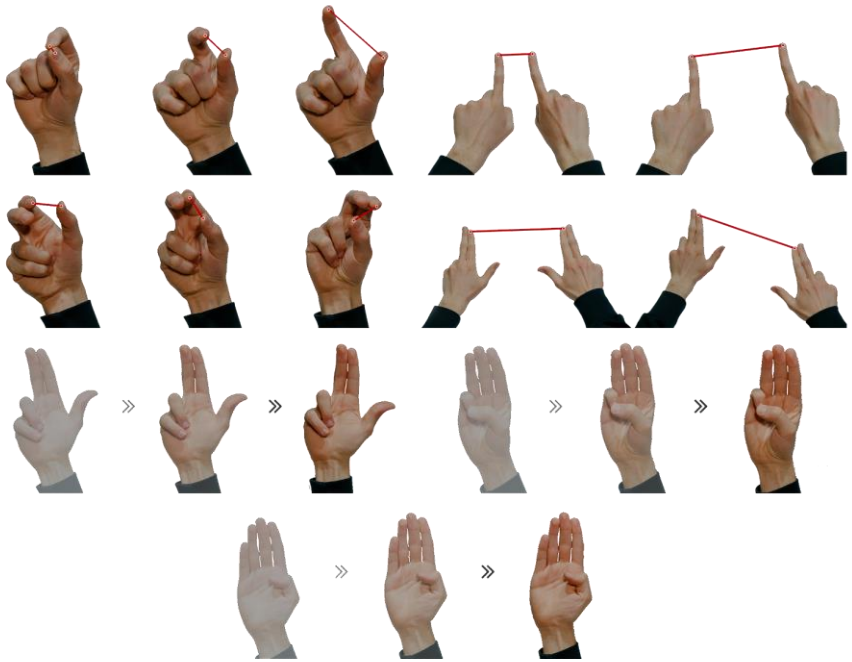
System pipeline



Dataset



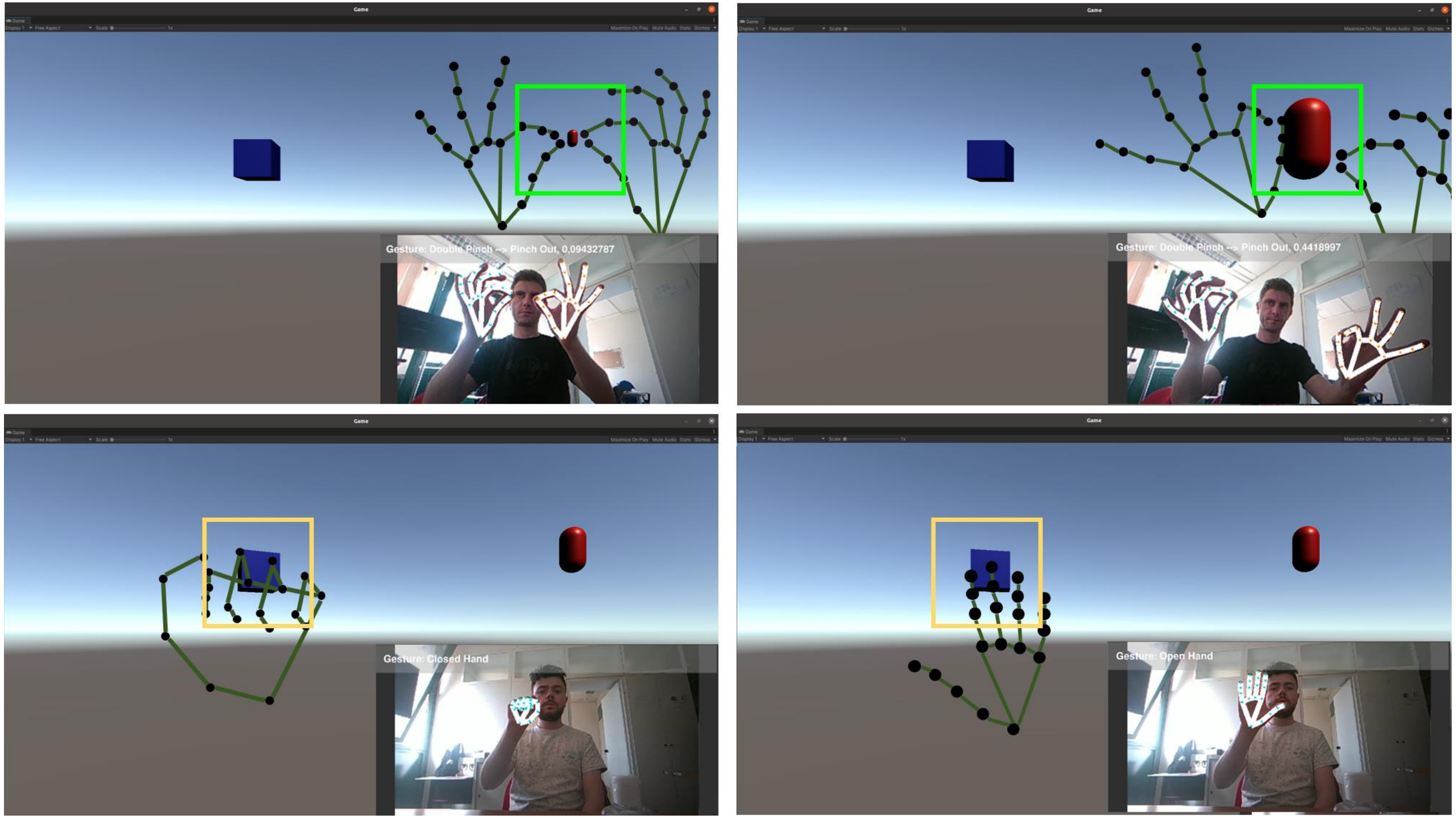
Static Gestures



Dynamic gestures

XR Application

The gesture recognition system presented was employed in a collaborative augmented reality application for real-time 3D scene creation.



Publication

*Nicola Capece, Gilda Manfredi, Vincenzo Macellaro, and Pietro Carratù. **An easy Hand Gesture Recognition System for XR-based collaborative purposes**. In IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering. IEEE, 2022.*

2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE) | 978-1-6654-8574-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/MetroXRAINE54828.2022.9967592

An easy Hand Gesture Recognition System for XR-based collaborative purposes

1st Nicola Capece, 2nd Gilda Manfredi, 3rd Vincenzo Macellaro
Department of Mathematics, Computer Science, and Economics
University of Basilicata, Potenza, Italy
nicola.capece@unibas.it, gilda.manfredi@unibas.it, vincenzo.macellaro@studenti.unibas.it

4th Pietro Carratù
Youbiquo S.r.l.
Cava de' Tirreni, Italy
pietro.carratu@youbiquo.eu

***Abstract*—In the last few years, collaborative tools have increased dramatically, partly to compensate for the distance between people due to the pandemic and partly to allow activities (work, entertainment, free time, *etc.*) to be carried out among people without having to worry about geographical distances. In this scenario, it was necessary to overcome the classic remote meeting tools (*e.g.* video, audio, chat), which have a reduced sense of presence. Extended Reality (XR) represents a Computer Graphics (CG) based innovative technology particularly suited to this purpose. Indeed, XR aims to develop Virtual Reality (VR), Augmented Reality (AR), and Mixed Reality (XR) solutions that can transform how people interact by increasing the sense of presence. This last aspect depends not only on the virtual/real objects and scene visualization but also on their interaction. In this regard, human-computer interaction (HCI) techniques represent a possible solution. However, these techniques depend on specific devices, such as Head Mounted Display (HMD), Smart Glasses, Depth and Tracking Cameras, *etc.*, whose costs make access difficult. For this reason, we propose a Hand Gesture Recognition (HGR) system that can be used in XR applications using a simple RGB camera. Our is a deep learning system based on MediaPipe, the state-of-the-art (SOTA) for hand tracking through simple RGB images [1], [2].**

***Index Terms*—Hand Gestures, Deep Learning, Human-computer Interaction**

I. Introduction

The need to reduce distances among people to ensure a high sense of presence has become one of the biggest challenges in computer science and CG. In the last years, more and more collaborative tools such as Mozilla Hubs, VRChat, and other general purposes browser-based web 3D tools [3] were developed and made accessible for users. Some of them ensure user presence by allowing the creation of personalized avatars [4], [5]. Generally, these tools guarantee high performance in terms of a sense of presence when scene visualization and interaction are performed using HMDs and their controllers. However, more recent HMDs are suitable for being combined with hand tracking devices, such as the Leap Motion controller, to allow interaction using freehand [6] and gesture recognition [7]. Furthermore, the last HMDs, such as Oculus Quest 2 and Vive Focus 3, integrate on-board hand-tracking sensors, capturing hand gestures directly and ensuring greater ease of handling and ergonomics, increasing the sense of presence, body ownership, embodiment, and agency [5], [8] of the users. As these devices are not accessible to all users due to their cost and usability, new technologies have recently emerged that allow freehand and gesture recognition through general-purpose devices. In this context, we propose a deep learning approach that allows HGR to use a simple, low-cost RGB camera.

In particular, we defined a system based on a well-structured pipeline in which the landmarks predicted from the MediaPipe Hands solution [2] are used as input to a simple feed-forward neural network (FFNN). The main advantage of our approach is that MediaPipe needs only a simple RGB hands-content camera frame as input to predict hands-landmarks. They can be obtained in real-time and are entirely independent of the camera features. The predicted landmarks correspond to the hand landmarks from which our FFNN can predict the corresponding hand gesture. To train our FFNN, we defined a dataset based on a dictionary of 15 gestures represented by many combinations of open and closed fingers and hands. Gestures are divided into static and dynamic based on their behavior. The former consisted of the FFNN predicted gesture, and the latter were obtained using the FFNN prediction as an activation gesture and performing assertion on the hands-landmarks for the subsequent frames.

We used our system for collaborative real-time 3D scene interaction in the XR environment using the well-known Unity 3D game engine and its network library called Netcode [1]. The proposed 3D application is a simple authoring tool use case helpful in building 3D scenes in real-time through collaborative multi-user interaction. Since our system is trained with landmarks captured in the egocentric mode [9], it can be used with smart glasses as an on-device system.

The remainder of this paper is structured as follows: in Section II an overview of related work was provided; in Section III we provide a background and motivations; in Section IV we described our proposed HGR system; A XR collaborative authoring use case is described in Section V, and finally the conclusions suggestions for future works were provided in Section VI.

II. Related Work

The 3D interaction with the hands benefits significantly from using the HGR system [10]. The first step to having a working HGR system is to choose the correct data acquisition

[1] https://docs-multiplayer.unity3d.com/netcode/current/about

978-1-6654-8574-6/22/$31.00 ©2022 IEEE

Virtual Dressing Rooms

Virtual Dressing Rooms (VDRs) are an emerging technology that allows users to try on clothing virtually without physically wearing the clothes.

Computer vision and deep learning track user's body and simulate clothing in real-time.

Potential to revolutionize retail, providing personalized shopping and reducing physical inventory.

Our VDR solution

Purpose: overcome the limitations of virtual dressing rooms currently available on the market.

Source: https://style.me/

No clothing animations

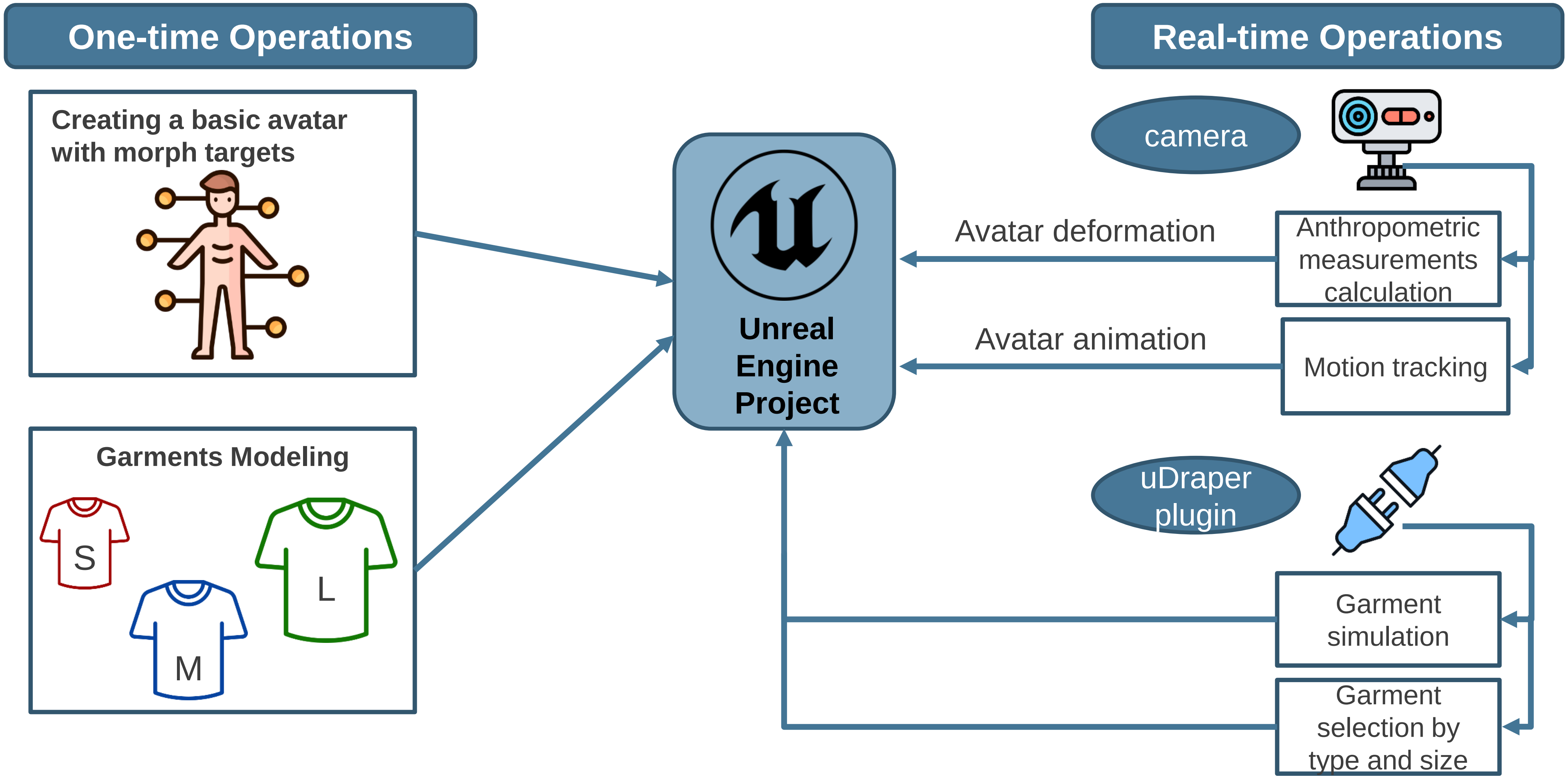
Source: asizer.com

Clothes that do not fit the user's body

Source: FXMirror 3D

Presence of artifacts in clothing

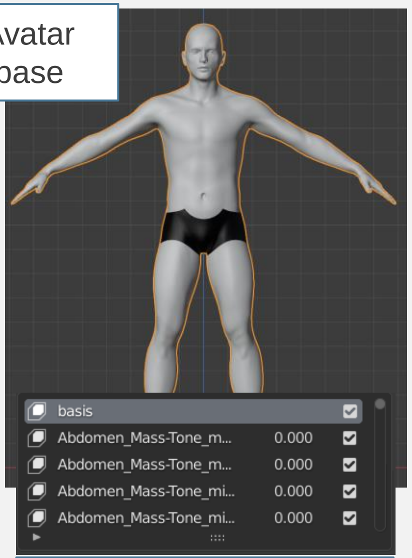
System Architecture



Phases of the basic avatar creation algorithm

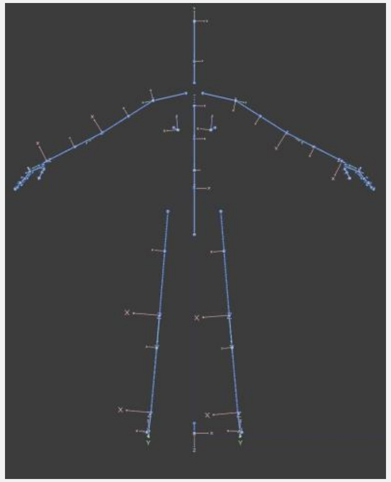
Phase 1: Creation of Mesh and Morph Targets

Avatar base

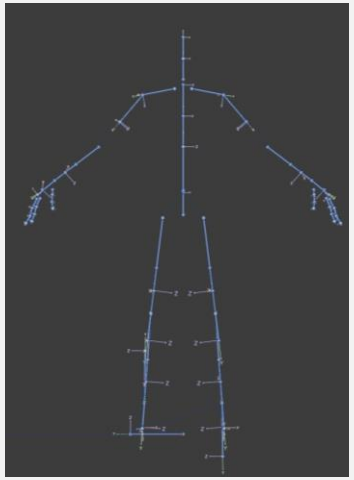


Morph targets

Phase 2: adaptation of the basic skeleton with the UE4 mannequin one

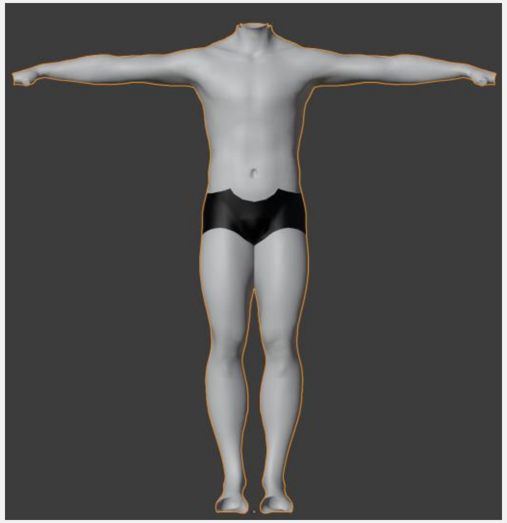


Base skeleton



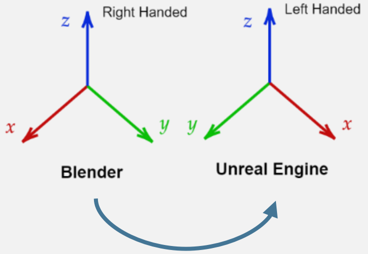
UE4 skeleton

Phase 3: creation of the low resolution Mesh for the clothing simulation engine



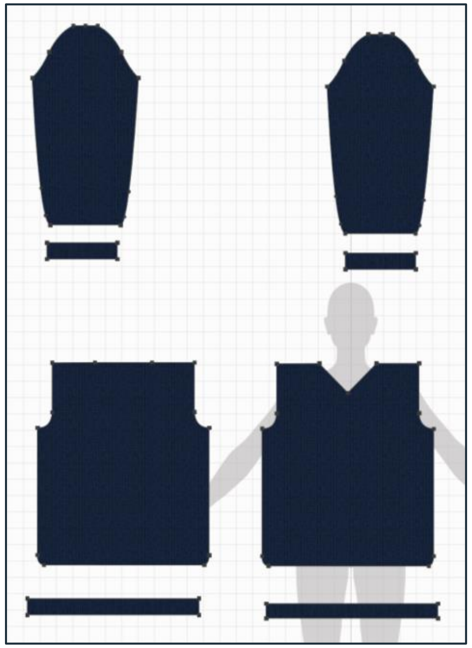
Step 4: Export as FBX file

- High-resolution mesh
- Low-resolution mesh
- skeleton
- morph targets
- materials

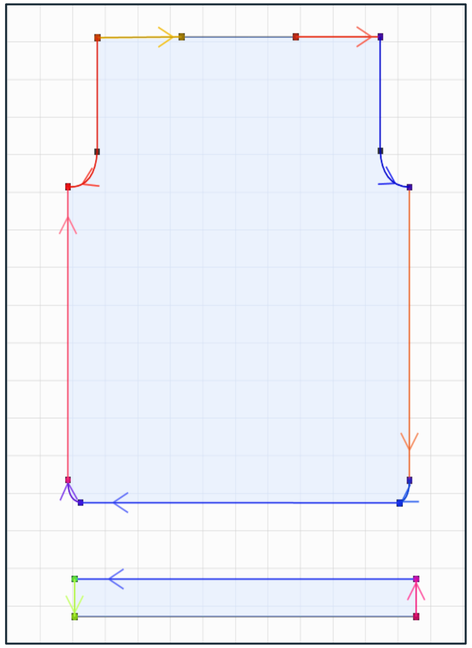


Garment modeling

The modeling process begins with a 2D paper pattern



Define the seams on the edges of the pattern



The pattern sections should be placed close to the base avatar

You have to set the physical parameters of the fabric

Cotton
thickness = …
damping = …

Dress simulation can be performed

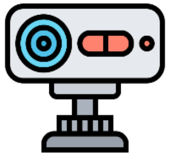
Camera Operations

Camera Plugin

Anthropometric measurements calculation

User's motion tracking

Camera Operations



Camera Plugin

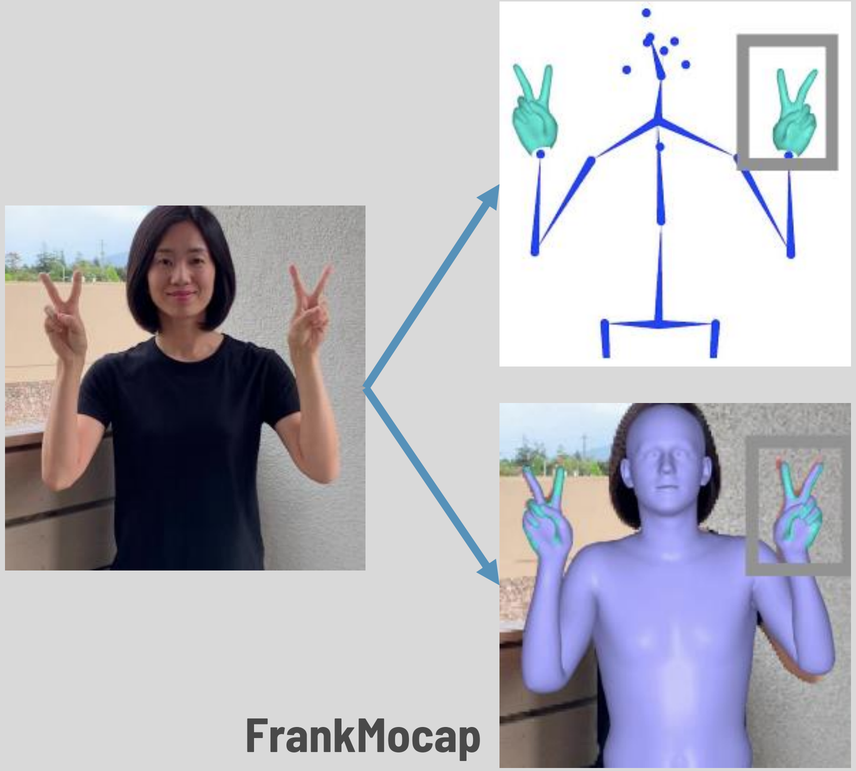
Rong, Yu et al. "FrankMocap: A Monocular 3D Whole-Body Pose Estimation System via Regression and Integration." 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW) (2021): 1749-1759.

The calculation of the anthropometric measures for the deformation of the avatar is an example of **procedural modeling**.

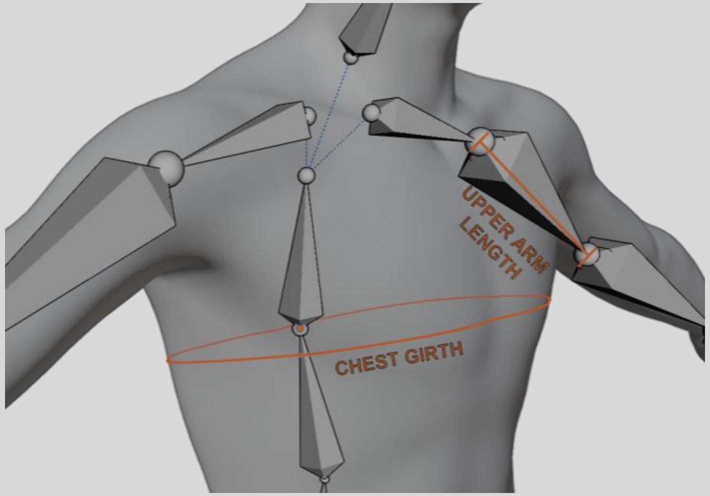
Anthropometric measurements calculation

Application for calculating anthropometric measurements

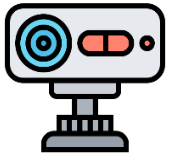
We use skeleton and mesh reconstructed by FrankMocap to calculate the anthropometric measurements



FrankMocap



Camera Operations



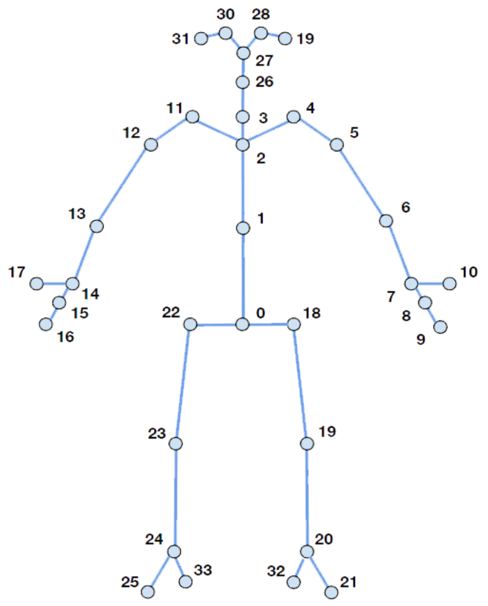
Camera Plugin

User’s motion tracking

While the calculation of anthropometric measurements is performed only once for each user, the tracking of movements takes place for each camera frame.

ZED camera

Uses Neural Networks to get 2D and 3D information (position and orientation) of a series of body keypoints.



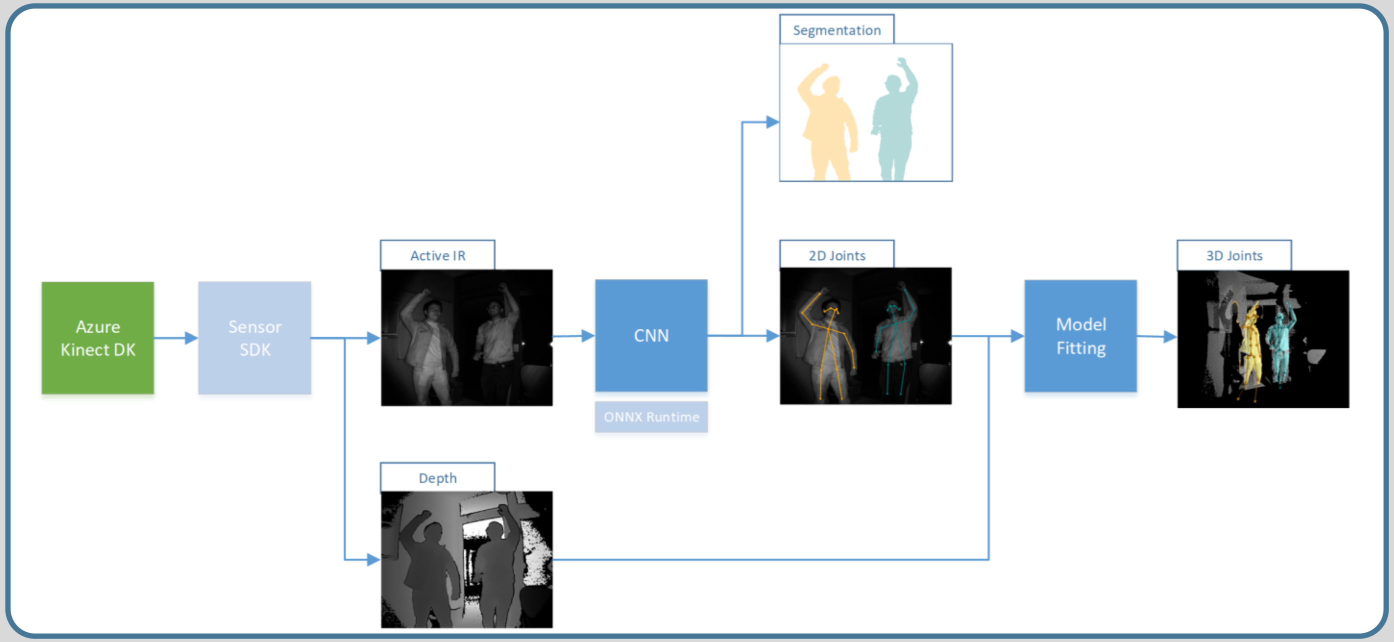
Camera Operations

Camera Plugin

User's motion tracking

While the calculation of anthropometric measurements is performed only once for each user, the tracking of movements takes place for each camera frame.

Azure Kinect camera



Camera Operations

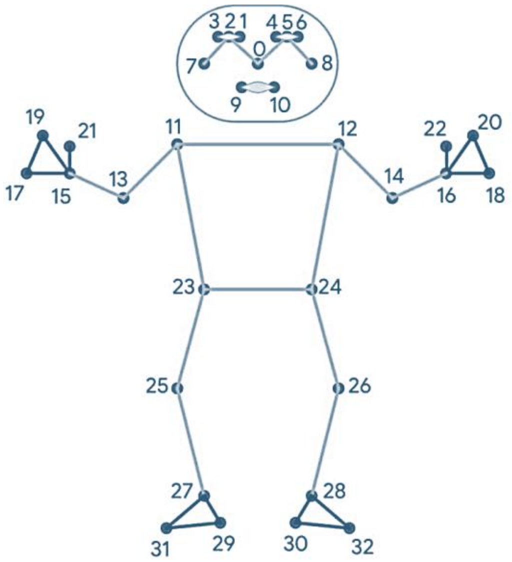
Camera Plugin

User’s motion tracking

While the calculation of anthropometric measurements is performed only once for each user, the tracking of movements takes place for each camera frame.

Simple RGB camera

- Utilizes MediaPipe’s convolutional neural network for joint detection and tracking
- Trained on large labeled image datasets for accuracy
- Estimates 3D positions of detected joints
- The body tracking plugin maps joints to UE4 mannequin skeleton for animation



Publication

Gilda Manfredi, Nicola Capece, Ugo Erra, Gabriele Gilio, Vincenzo Baldi, and Simone Gerardo Di Domenico. ***TryItOn: A Virtual Dressing Room with Motion Tracking and Physically Based Garment Simulation****. In International Conference on Extended Reality, pages 63–76. Springer, 2022*

Check for updates

TryItOn: A Virtual Dressing Room with Motion Tracking and Physically Based Garment Simulation

Gilda Manfredi[1](✉), Nicola Capece[1], Ugo Erra[1], Gabriele Gilio[1], Vincenzo Baldi[2], and Simone Gerardo Di Domenico[2]

[1] University of Basilicata, 85100 Potenza, PZ, Italy
{gilda.manfredi,nicola.capece,ugo.erra,gabriele.gilio}@unibas.it
[2] SAUTECH S.R.L., 84013 Cava de' Tirreni, SA, Italy
{vincenzo.baldi,simone.didomenico}@sautechgroup.com

Abstract. The interest in the online shopping field has increased considerably nowadays. This has led to emerging technologies being applied in this context. One of the sectors to which particular attention has been paid is the clothing sale. The main reason comes from the problem of physically trying on clothes to choose the correct size. This issue can be addressed using methodologies based on virtual reality and Human-Computer Interaction (HCI). This paper presents a Virtual Dressing Room (VDR) application named TryItOn that allows the user to try on digital clothes. The proposed solution offers an immersive and realistic experience by combining a high degree of photorealism, the presence of an avatar with accurate measurements, the natural interaction with the environment through movements of hands or the entire body, and the use of a real-time physical simulation of the garment.

Keywords: Virtual Dressing Room · Body tracking · Garment physical simulation

1 Introduction

In the last few years, the entire world population has faced a serious epidemic that has led us to gradually change our lifestyles and habits. Fear of contagion has shown people to reduce unnecessary travel and to pay more attention to hygiene. This change is also reflected in the purchase mode of goods such as clothing and footwear. In this regard, many studies conducted in various parts of the world show that there has been an increase in purchases made using e-commerce platforms since the beginning of the pandemic. To confirm this thesis, we conducted a study on consumers' buying habits, which showed that, despite the increase in online purchases, many people still prefer to shop in-store if clothing items have to be purchased. The customer can try on different sizes in the shop and choose the correct one, an action impossible to perform on a standard website for online shopping. As a demonstration of this, it is possible

L. T. De Paolis et al. (Eds.): XR Salento 2022, LNCS 13445, pp. 63–76, 2022.
https://doi.org/10.1007/978-3-031-15546-8_5

Thank you for your attention