

Introduction

Telepresence on Egocentric Fisheye Camera

Hardware Setup Egocentric Images Free-viewpoint Avatars

We propose EgoRenderer, a mobile end-to-end telepresence system.

- Hardware setup: A wearable fisheye camera mounted on a cap or a VR headset.
- Input: Real-time captured egocentric sequences.
- Output: Free-viewpoint full-body avatars.

EgoRenderer: A Mobile Telepresence System

Pipeline:

- Texture Synthesis by real-time egocentric texture transfer and implicit texture learning
- Pose Construction
- Neural Rendering

Challenges of the egocentric setup:

- Large distortions, top-down view, and self-occlusions of egocentric images
- Lack of ground truth for real egocentric images

The pipeline consists of the following steps:

- Ego-Image** is processed by **Ego-DPNet** to produce **Ego-Pose**.
- Ego-Image** and **Ego-Pose** are used for **UV Texture Extraction** and **Pose Estimation**.
- Pose Estimation** leads to **3D Joint Pose**, which is processed by **Inverse Kinematics** to create a **3D Model**.
- The **3D Model** is projected into **Target-Pose in UV space**.
- Texture Synthesis** involves **Texture Transfer** and **Texture Latents** from the **Global Texture-Stack**.
- Neural Rendering** uses **Feature Image** and **RenderNet** to produce the **Generated Avatar**.
- Losses used for training include **Adversarial Loss**, **Face Identity Loss**, and **VGG Loss**.

Experiments & Results

Compared with DNR, the texture transfer of Ego-Img leads to improvements in visible regions in the Ego-Image (e.g., face, neck and collar) and we even restore the tiny buttons.

Our method, based on dynamic and implicit textures, can generate time- and pose-dependent appearances

LPIPS [Zhang et al.] Scores of Each Method

Method	LPIPS Score
FeaNet [Sarkar et al.]	1.54
DNR [Thies et al.]	1.58
Pix2PixHD [Wang et al.]	1.62
Ours	1.66

Compared with existing methods, our method synthesize avatars with smaller LPIPS

Contributions

- A large synthetic training dataset of egocentric fisheye images, and an Ego-DPNet network to predict dense correspondence.
- An end-to-end mobile telepresence system that takes single egocentric images as input and generates free-viewpoint full-body avatars for egocentric setup.