A Video-Texture Based Approach for Realistic Avatars

of Co-Located Users in Immersive Virtual Environments Using Low-Cost Hardware

Authors

Robin Horst,

Sebastian Alberternst,

Jan Sutter,

Philipp Slusallek,

Uwe Kloos,

Ralf Dörner

Conference

International Conference on Computer Graphics Theory and Applications

Year: 2019

Problem:

Realism for interaction between users inside a virtual environment.

- Collaboration
- Accuracy
- Body Language
- Preventing effects such as Proteus Effect

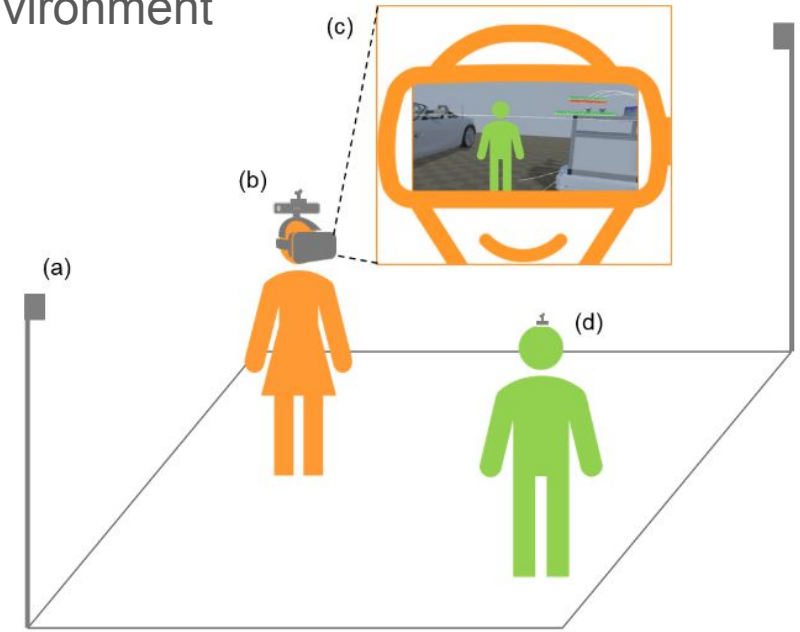
Actual Solutions

- Highly realistic avatars.
- Customization
- Trackers & IA
- Perspective information

Paper’s solution

RGB capture from point of view

Add user in a 2D plane in the Virtual Environment



Paper’s solution

- Low cost setup
- Occlusion techniques
- Image segmentation

Related Works

Other works explore similar approaches:

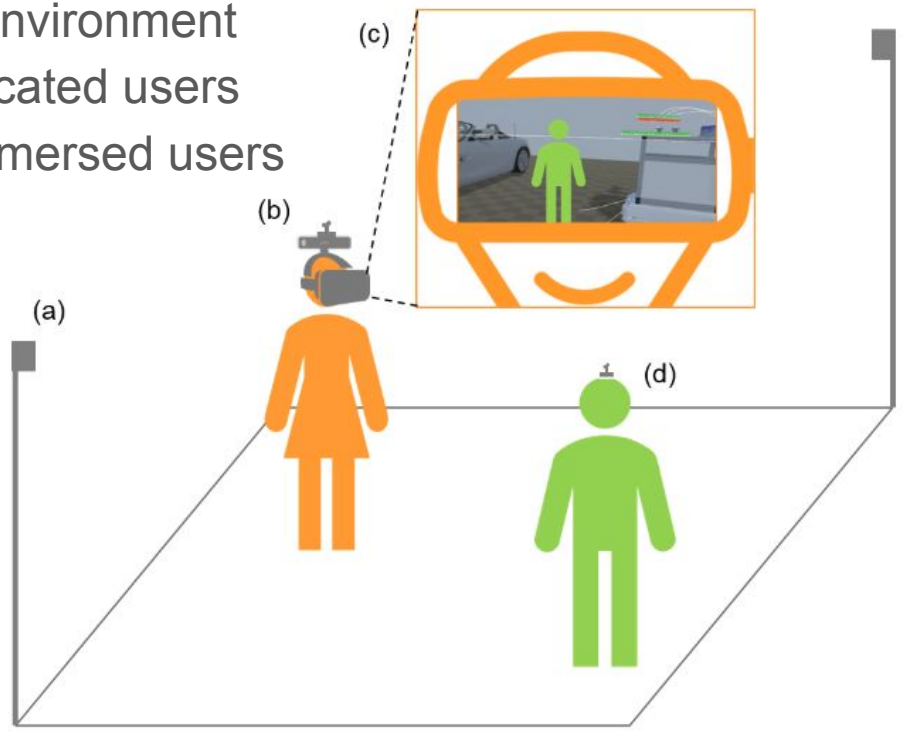
- Realistic 3D models
- Realistic tracking techniques
- Depth information
- Chroma Screens

LOW-COST VIDEO-TEXTURE APPROACH

1. Technical Setup
2. Capturing People
3. Situating People
4. User Study
5. Evaluation

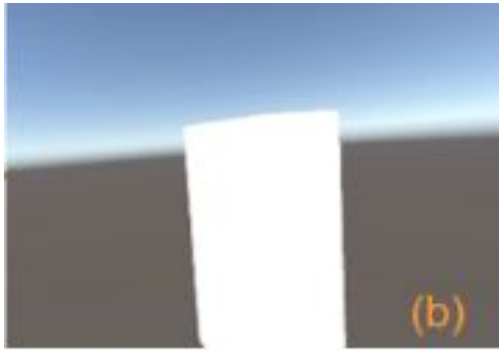
LOW-COST VIDEO-TEXTURE: Technical Setup

- Room with the size of the VR environment
- tracker for immersed and co-located users
- RGB and Depth sensors for immersed users



LOW-COST VIDEO-TEXTURE: Technical Setup

- RGB
- Depth
- User position
- User Prespective

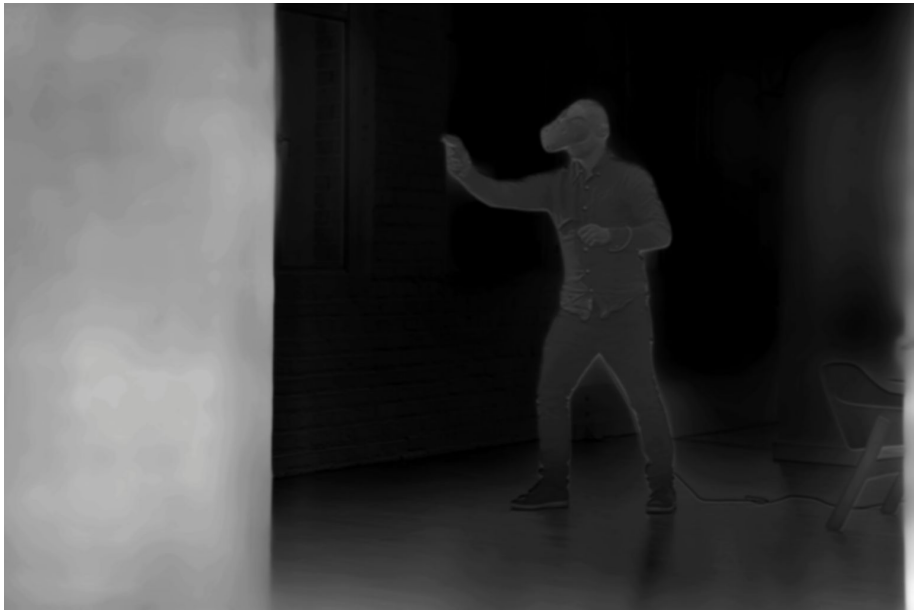


LOW-COST VIDEO-TEXTURE: Capturing People

- BGRA image of size 1920x1080

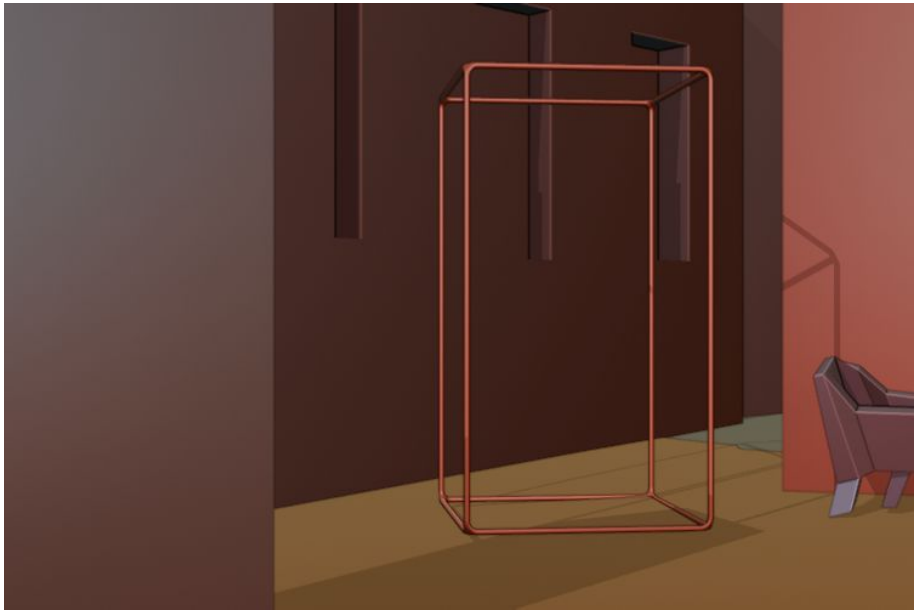
LOW-COST VIDEO-TEXTURE: Capturing People

- Depth image of size 512x424



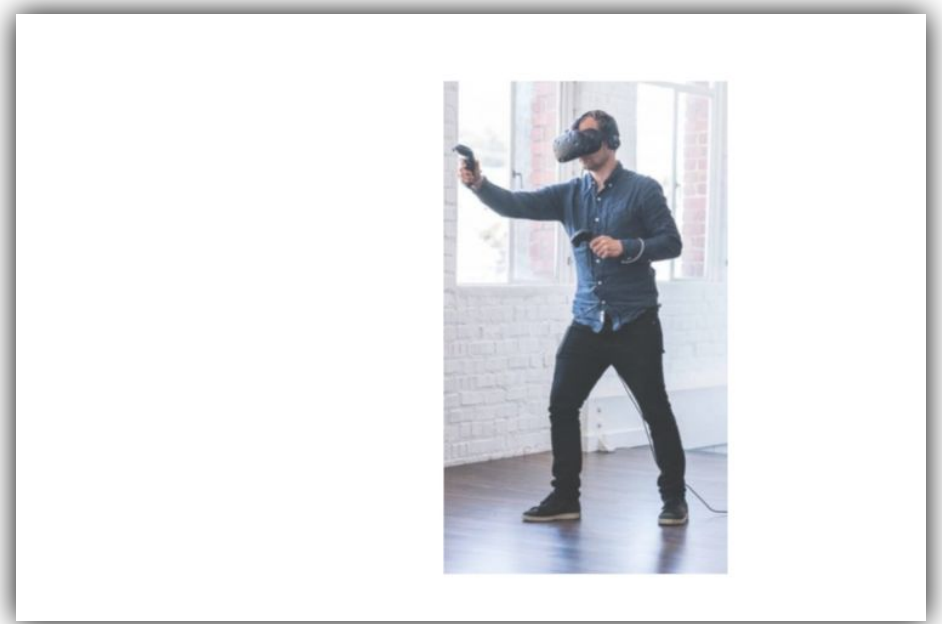
LOW-COST VIDEO-TEXTURE: Capturing People

- Position of the user



LOW-COST VIDEO-TEXTURE: Capturing People

- Bounding Box as a mask



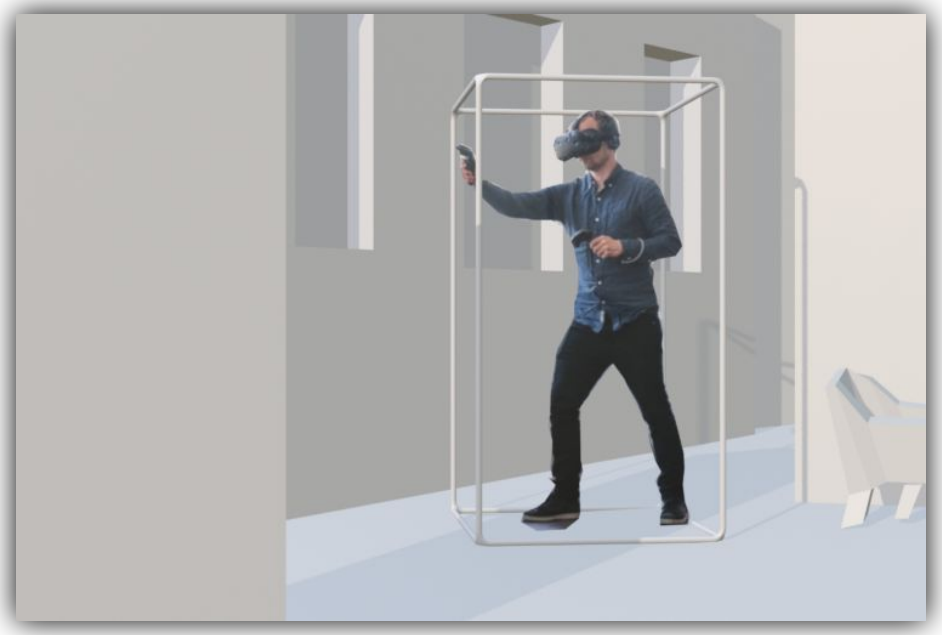
LOW-COST VIDEO-TEXTURE: Capturing People

- Depth + bounding box position as mask

$BXmin \leq X \leq BXmax$

$BYmin \leq Y \leq BYmax$

$BZmin \leq Z \leq BZmax$



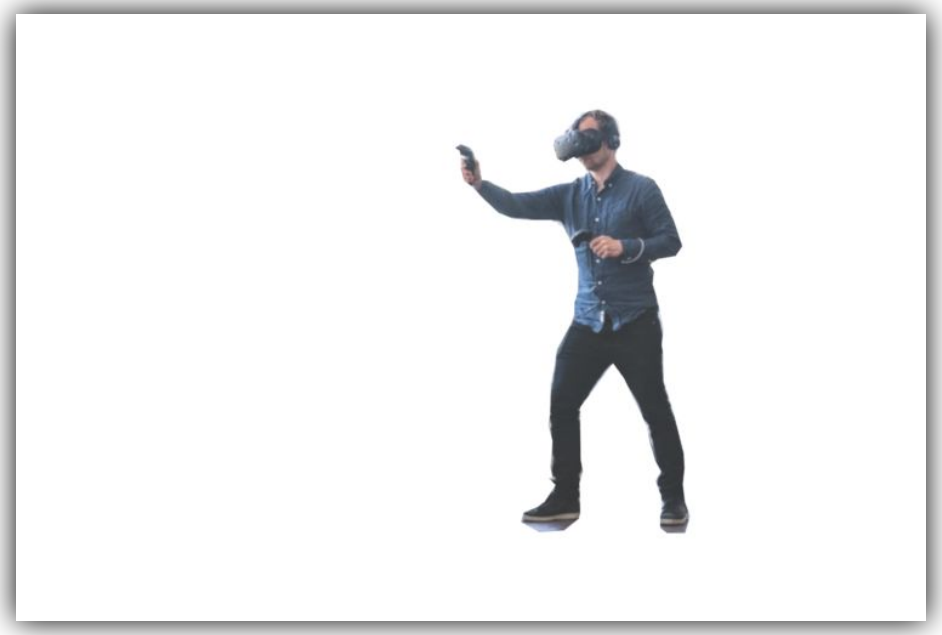
LOW-COST VIDEO-TEXTURE: Capturing People

- Depth + bounding box position as mask

$BX_{min} \leq X \leq BX_{max}$

$BY_{min} \leq Y \leq BY_{max}$

$BZ_{min} \leq Z \leq BZ_{max}$



LOW-COST VIDEO-TEXTURE: Capturing People

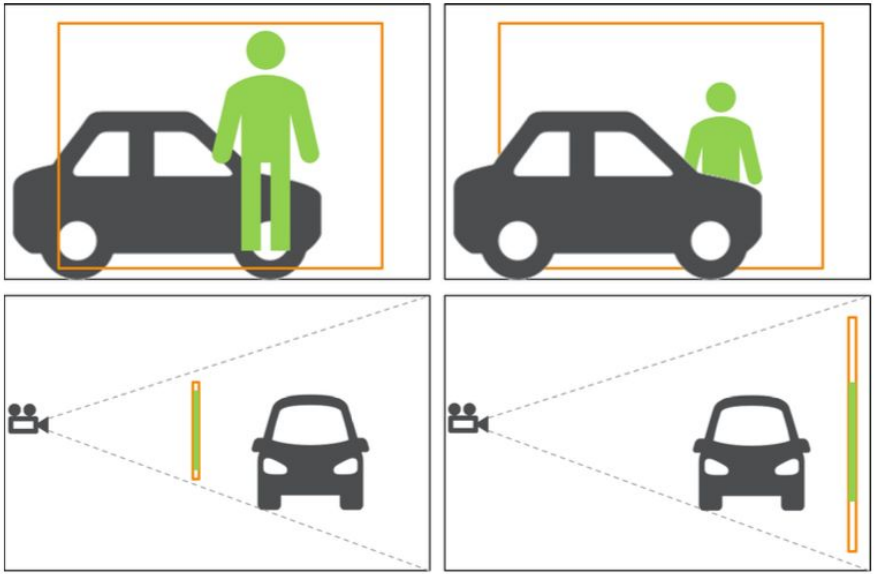
Papers Pseducode:

Algorithm 1 Algorithm for VR bounding box segmentation of one image

INIT arrays: *KinectDepthImage, KinectColorImage, BB˙DepthImage, BB˙ColorImage, OutputImage*
INIT ints: *UserToUserDistance, BB˙Indices*
1: fetch data from different external sources
2: wait until every source is available
3: **for all** pixels in *KinectColorImage* **do**
4: **if** (pixeldepth of *KinectDepthImage* = invalid) **or** (pixelcolor of *BB˙ColorImage* ! = pure white) **or** (pixeldepth of *KinectDepthImage* > pixeldepth of *BB˙DepthImage*) **then**
5: write alpha value 0 in *OutputImage* pixel
6: **else if** (*UserToUserDistance* >= pixeldepth in *KinectDepthImage* > 0) **then**
7: write pixelcolor of *KinectColorImage* in *OutputImage* pixel and update *BB˙Indices*
8: **else**
9: write alpha value 0 in *OutputImage* pixel
10: **end if**
11: **end for**
12: **if** (*BB˙Indices* describe a valid bounding box) **then**
13: send sub-image of *OutputImage* and *BB˙Indices* to VE instance for further texture integration
14: **end if**

LOW-COST VIDEO-TEXTURE: Situating People

- Video texture is drawn on a billboard
- Distance to camera adjusted according to tracker data
- Absolute size adjusted to compensate for distance



Occlusion handling

- Mapping 3D data onto a 2D plane causes a loss of a dimension
- In certain situations the co-located avatar should occlude virtual objects but doesn’t (with outreaching arms)
- To compensate for this the virtual avatar is offset towards the virtual camera by 70cm, projecting it at roughly the edge of the bounding box

Evaluation: User test design

- 15 participants between the ages of 23 and 35
- 10 male, 5 female
- Recruited from the university environment and engineering staff

Procedure:

- Demographic questionnaire
- Technical setup introduction
- 12 minutes of interaction with experimenter
- Questionnaire with open ended questions

Evaluation: Findings

- Video texture avatars were useful and adequate
- Non verbal cues could be interpreted well at close distances

Issues:

- On some angles parts of the avatar were cut off due to bad coverage by kinect
- At larger distances the quality of the avatar diminished and it became harder to interpret
- The camera rig was uncomfortable
- Lighting of the avatar was highly unnatural

Conclusions by Authors

- Low cost hardware solution
- No dependence on pre trained computer vision
- No prior 3D scanning
- No texture baking
- View independent rendering
- Focus on co-located people

Our Conclusions

- Hardware is being integrated into commercial headsets e.g. Quest 3
  - This solves the problem with clunky hat
- Good use of limited resources for real time avatars
- More work is required for integrating real life and virtual lighting
  - Could take clues from differential rendering
  - Dynamic lightning conditions on the VR world should affect the avatar
  - Real world lightning conditions have to be taken into account carefully adding complexity to the setup