

Real-Time 3D volumetric human body reconstruction from single view RGB-D capture device

Rafael Diniz and [Mylène C.Q. Farias](#)

University of Brasília, Brazil

<http://www.ene.unb.br/mylene>

El-3DMP, January 16, 2019, California



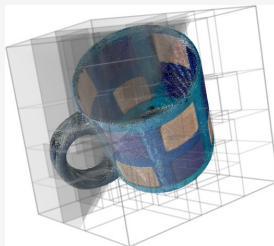
Universidade de Brasília

Summary

- Motivation and Goals
- Review of current work
- Proposed work
- Preliminary Results
- Conclusions and Future Work

Volumetric video

- High popularity of immersive experiences;
- **Point Clouds:**
 - Set of coordinates indicating the location of each point, along with one or more attributes such as color associated with each point;
 - Alternative 3D content representation that allows visualization of scenes in a more immersive way;
 - Viable solution to represent visual stimuli because of the efficiency and simplicity for capturing, storing and rendering of 3D objects;



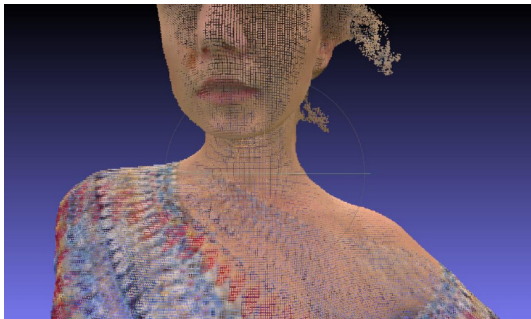
Challenges

- 3D point clouds are captured using multiple cameras and depth sensors in various setups;
- This results in thousands up to billions of points in order to represent realistically reconstructed objects or scenes;



Challenges

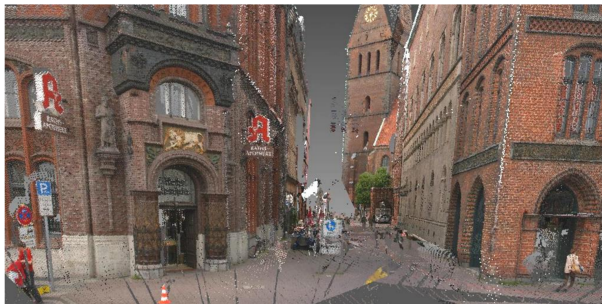
- Each point may have a 3D position information (x, y, z), a color information (R, G, B), and possibly attributes such like transparency, time of acquisition, reflectance of laser, etc.
- Efficient representation of point clouds are needed to store or transmit these information;
- Compression is much more difficult because each point is basically not related each other, e.g., no orders and no local topology exists.



Motivation

Applications

- It is believed that a wide range of applications and use cases can benefit from this type of data representation;



02/01/2017

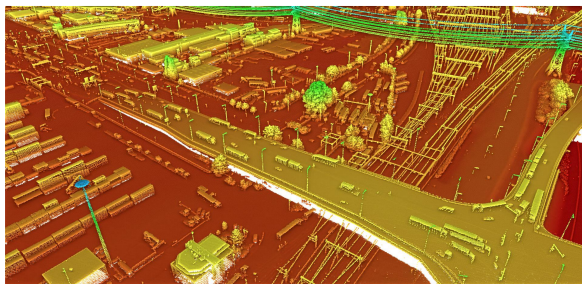
Capturing Reality with Point Clouds

Copyright © 2017 Hasso Plattner Institute / Rico Richter. All rights reserved.

Motivation

Applications

- It is believed that a wide range of applications and use cases can benefit from this type of data representation;



02/01/2017

Capturing Reality with Point Clouds

Copyright © 2017 Hasso Plattner Institute / Rico Richter. All rights reserved.

9

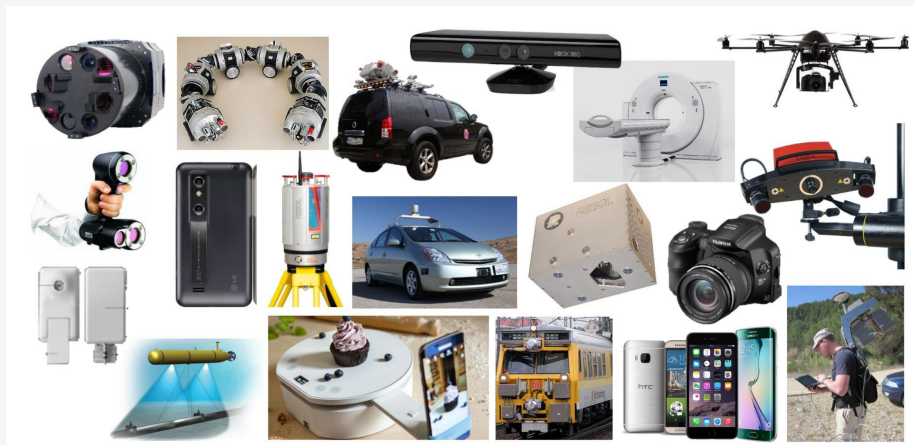
Motivation

Applications

- Video communications is certainly one exciting application, but to be popular the equipment must be affordable to the final regular user.



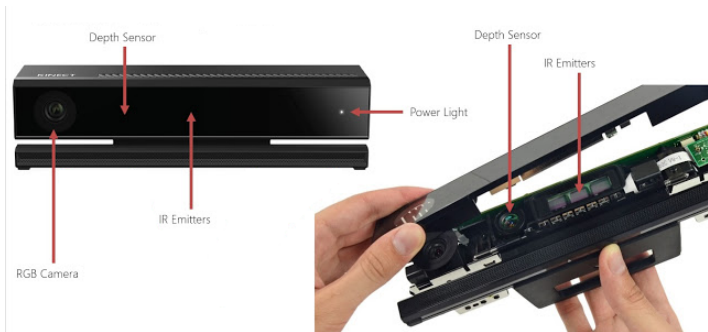
Motivation



Capturing

Kinect-2 capture device

- Time-of-Flight depth ranging technology
- Depth resolution of 512x424, distances from 0.5 to 4.5m, FoV of 70.6° by 60° (HxV), millimeter accuracy
- RGB in HD, downconverted to match depth resolution



Typical volumetric video capture setup

Example with 4 Kinect 2 devices.



Proposed Work

Goals:

- Design a simple and affordable human body reconstruction system, which reconstructs a 3d person representation from a live input from a single RGB-D camera and a pre-captured 3D model;
- Use 1 Kinetic-2 device and a simple multicore CPU, making this the system affordable for regular users;
- Main target applications are **video telepresence communication**.

Methodology

Experimental Setup

- Code written in C and C++;
- Libraries used: Open3D and Libfreenect;
- Kinect-2 as main capture device;
- Intel Xeon E5-2620, with 80GB of RAM hardware;
- Lenovo T430 for mobile capture device.



Methodology

Constraints:

- Assumes that the back of the head of the person is rigid;
- The speaker is looking ahead during most of the time;
- Self-occlusions do not occur often;
- This way, higher dynamics of the object (mouth, nose, eyes) can be fully present in the reconstructed 3D volumetric stream;
- The method can be extended to other types of objects.

Methodology

Step 1

Create a volumetric representation of each person joining the volumetric video session:

- Use only 1 kinetic-2 device to capture the model by moving the capture device around the the person;
- Use Truncated Signed Distance Function and Kinect Fusion to assemble the volumetric 3D object representation.



Captured model example

Methodology

Step 1



Captured model examples

Methodology

Step 2

Then model is segmented and stored:

- The segmentation method uses RGB and depth;
- The nose information (maximum or minimum depth, depending on coordinate system) and its neighboring region, including the eyes, are segmented separately;
- The back of the head is segmented separately, using similar approach;
- This optimizes the efficiency of the registration between the model and the input live volumetric stream, by using smaller point-cloud inputs.



Segmented models

Methodology

Step 2



Segmented model examples

Methodology

Step 3

After the model is captured, the live capture system can start, and a pre-processing step is done for each captured RGB and Depth frame pair:

- For each RGB and Depth frame pair captured, an alignment is necessary because the timestamps of the color and depth frames differ between 10ms to 20ms, which although less than a frame period ($\sim 33\text{ms}$ at 30fps), is important specially for high speed movements;
- The RGB and Depth frames are converted to point-cloud, with camera coordinates converted to world coordinates by the use of Kinect's intrinsics parameters;



Point-cloud created from a live RGB-D input

Methodology

Step 3



Point-cloud from live single view capture examples

Step 4

We perform proposed volumetric object reconstruction from a single RGB-D sensor:

- In a similar approach of [Step 2](#), the point-cloud obtained from the live feed has its nose and adjacency area segmented;
- Using a fast global registration method between the segmented face from the model in [Step 2](#) and the segmented input point-cloud, a transformation matrix is obtained;
- The back of the head segmented from the model in [Step 2](#) is transformed using the transformation matrix;
- Finally, the transformed 3D model and live captured point-cloud are merged and the live reconstructed volumetric video frame is created.



Reconstructed frame

Methodology

Step 4



Reconstructed Point-cloud

Results

- Realtime CPU execution - under 33ms at 30fps;
- Better experience when compared to incomplete objects;
- Allows volumetric video use cases using single RGB-D capture device;
- Room left for GPU offloading optimization;
- Can be extended to different type of objects;



Conclusions

- The pre-processing steps made into the point-clouds before the registration step are important for both quality and real-time execution target;
- A fast global registration is needed, at the cost of having sometimes an imperfect transformation matrix;
- Work shows Mixed Reality Volumetric Video use cases using single RGB-D camera running in an affordable CPU is possible;
- Proposed framework can be further extended to different type of objects, especially the ones with large rigid areas.

Issues to be addressed

- Solve the color difference problems between model and input frame;
- Fusion / Merge of Point Cloud issues;
- Improve segmentation using Machine Learning

mylene@ene.unb.br,
http://www.ene.unb.br/mylene
http://www.ene.unb.br/mylene/databases.html