INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG1
CODING OF STILL PICTURES
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11 m38673
ISO/IEC JTC1/SC29/WG1 M72012

May 2016, Geneva, Switzerland

Source	**Microsoft**
Status	**Information for Joint Ad hoc Group for digital representations of light/sound fields for immersive media applications**
Title	**Microsoft Voxelized Upper Bodies – A Voxelized Point Cloud Dataset**
Authors	**Charles Loop, Qin Cai, Sergio Orts Escolano, and Philip A. Chou**

1. Introduction

Microsoft hereby makes available dynamic voxelized point cloud data sequences as potential test material for MPEG and/or JPEG standardization efforts, as well as non-commercial use subject to the accompanying license agreement by the wider research community.

A *voxelized point cloud* is a set of points (x, y, z) constrained to lie on a regular 3D grid, which without loss of generality may be assume to be the integer lattice. The (x, y, z) coordinates may be interpreted as the address of a volumetric element, or *voxel*. A voxel whose address is in the set is said to be *occupied*; otherwise it is *unoccupied*. Each occupied voxel may have *attributes*, such as color, transparency, normals, curvature, and specularity. A voxelized point cloud captured at one instant of time is a *frame*. A *dynamic* voxelized point cloud is represented as a sequence of frames.

The dynamic voxelized point cloud sequences in this dataset are known as the Microsoft Voxelized Upper Bodies (MVUB). There are five subjects in the dataset, known as Andrew, David, Phil, Ricardo, and Sara, pictured below. The upper bodies of these subjects are captured by four frontal RGBD cameras, at 30 fps, over a 7-10 s period for each. Two spatial resolutions are provided for each sequence: a cube of 512x512x512 voxels and a cube of 1024x1024x1024 voxels, respectively known as depth 9 and depth 10. In each cube, only voxels near the surface of the subjects are occupied. The attributes of a voxel are the red, green, and blue components of the surface color.

Voxels at depth 9 are approximately 1.5 mm cubed, while voxels at depth 10 are approximately 0.75 mm cubed.


Andrew


David


Phil


Ricardo


Sarah

2. Availability

The dataset is initially made available at ftp://ftp.research.microsoft.com/users/pachou/MVUB/. The location of the data may be moved to MPEG and/or JPEG servers. The MVUB directory contains two files: MVUB.zip and this document. MVUB.zip is 10 GB. Unzipped, it is 40 GB.

3. Folder Structure

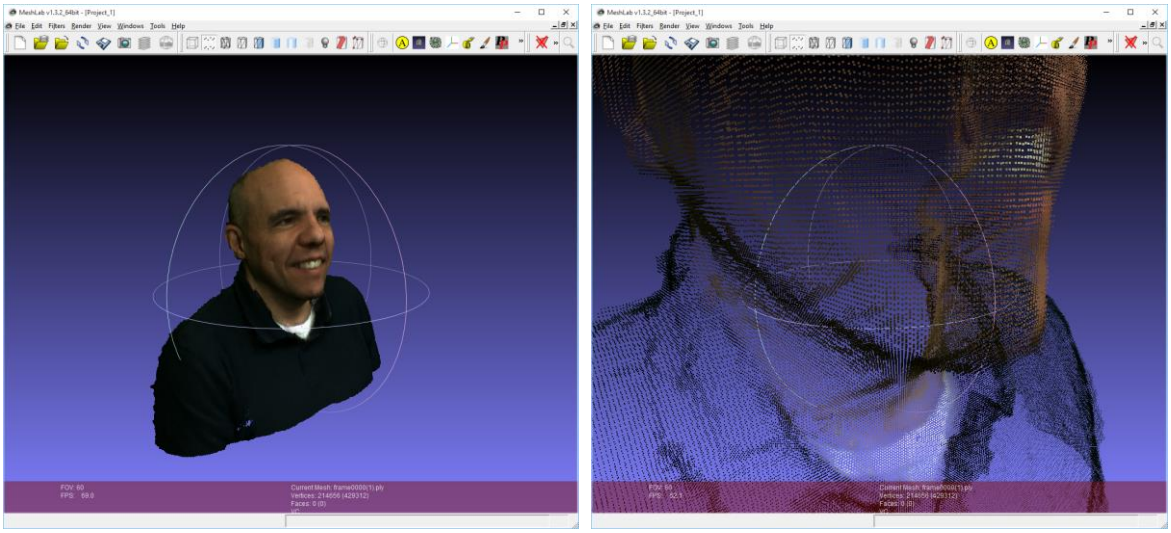
The structure of the unzipped folder is as follows:

- Ply/ – main data, as separate files for each sequence, depth, and frame
  - Andrew9/frame0000.ply, …, Andrew9/frame0317.ply
  - Andrew10/frame0000.ply, …, Andrew10/frame0317.ply
  - David9/frame0000.ply, …, David9/frame0215.ply
  - David10/frame0000.ply, …, David10/frame0215.ply
  - Phil9/frame0000.ply, …, Phil9/frame0244.ply
  - Phil10/frame0000.ply, …, Phil10/frame0244.ply
  - Ricardo9/frame0000.ply, …, Ricardo9/frame0215.ply
  - Ricardo10/frame0000.ply, …, Ricardo10/frame0215.ply
  - Sarah9/frame0000.ply, …, Sarah9/frame0206.ply
  - Sarah10/frame0000.ply, …, Sarah10/frame0206.ply
- Png/ – images from orthogonal frontal projection for each sequence, depth, and frame
  - Andrew9/frame0000.png, …, Andrew9/frame0317.png
  - Andrew10/frame0000.png, …, Andrew10/frame0317.png
  - David9/frame0000.png, …, David9/frame0215.png
  - David10/frame0000.png, …, David10/frame0215.png
  - Phil9/frame0000.png, …, Phil9/frame0244.png
  - Phil10/frame0000.png, …, Phil10/frame0244.png
  - Ricardo9/frame0000.png, …, Ricardo9/frame0215.png
  - Ricardo10/frame0000.png, …, Ricardo10/frame0215.png
  - Sarah9/frame0000.png, …, Sarah9/frame0206.png
  - Sarah10/frame0000.png, …, Sarah10/frame0206.png

- Avi/ – videos from png files, for each sequence and depth
    - Andrew9.avi, Andrew10.avi
    - David9.avi, David10.avi
    - Phil9.avi, Phil10.avi
    - Ricardo9.avi, Ricardo10.avi
    - Sarah9.avi, Sarah10.avi
- License.pdf – Microsoft Research License Agreement

4. File Format

The main data are in the PLY file format [1], and if desired may be viewed directly in software such as MeshLab [2], as seen below.



The beginning of Andrew9/frame0000.ply, for example, is:

```
ply
format ascii 1.0
element vertex 279664
property float x
property float y
property float z
property uchar red
property uchar green
property uchar blue
end_header
34 255 0 0 235 0
34 255 1 123 167 130
35 255 2 126 154 131
36 254 0 135 163 140
36 254 1 133 161 138
37 254 1 130 158 135
37 255 0 129 157 134
37 255 1 142 147 143
36 255 2 183 203 176
```

```
38 254 0 138 158 131
38 255 0 139 159 132
38 255 1 176 196 169
39 254 0 147 167 140
39 255 0 168 201 193
41 251 0 139 155 142
.
.
.
```

The X, Y, and Z coordinates are integers in the range 0, …, 511 for the depth 9 sequences, and in the range 0, …, 1023 for the depth 10 sequences. The R, G, and B coordinates are integers in the range 0, …, 255.

The coordinate system is right handed, with the cube containing the subject having its origin (0,0,0) ahead and to the lower right from the subject's point of view, the side of the cube with the maximum X-coordinate to the subject's left, the side of the cube with the maximum Y-coordinate behind the subject's back, and the side of the cube with the maximum Z-coordinate above the subject's head. That is, the X-axis points to the subject's left; the Y-axis points to the subject's rear, and the Z-axis points up.

5. Processing

The data are produced by a real time capture system with four synchronized RGBD inputs. The cameras are placed at the frontal, frontal low, left and right profile positions within 1 m to 1.5 m distance from the subjects. The depth inputs are fused into a 3D voxel model in real-time, using a hierarchical carving approach. An initial grid of large voxels is tested for possible surface containment. Voxels that cannot contain surface, i.e., free space or inside a solid, are culled. Voxels that might contain surface are subdivided into 8 sub-voxels. This process repeats until a desired depth is reached (9 or 10 for these data sets). Voxels are colored according to a weighted average of their projections into the RGB images, similar to the method described in Loop et al. [3].

6. Citation

If you publish images or report performance results of these data, we request that you cite this document as Charles Loop, Qin Cai, Sergio Orts Escolano, and Philip A. Chou, "Microsoft Voxelized Upper Bodies – A Voxelized Point Cloud Dataset," ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document m38673/M72012, Geneva, May 2016.

7. References

[1] "The Stanford 3D Scanning Repository," http://graphics.stanford.edu/data/3Dscanrep/.

[2] "MeshLab," http://meshlab.sourceforge.net/.

[3] Charles Loop, Cha Zhang, and Zhengyou Zhang, "Real-Time High-Resolution Sparse Voxelization with Application to Image-Based Modeling," High Performance Graphics, 2013.