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

Source	**8i Labs, Inc.**
Status	**Input to Ad hoc Groups on MPEG Point Cloud Compression and JPEG PLENO**
Title	**8i Voxelized Full Bodies, version 2 – A Voxelized Point Cloud Dataset**
Authors	**Eugene d'Eon, Bob Harrison, Taos Myers and Philip A. Chou**

1. Introduction

8i Labs, Inc. 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 8i Voxelized Full Bodies, version 2 (8iVFBv2). There are four sequences in the dataset, known as *longdress*, *loot*, *redandblack*, and *soldier*, pictured below. In each sequence, the full body of a human subject is captured by 42 RGB cameras configured in 14 clusters (each cluster acting as a logical RGBD camera), at 30 fps, over a 10 s period. One spatial resolution is provided for each sequence: a cube of 1024x1024x1024 voxels, known as depth 10. In each cube, only voxels near the surface of the subjects are occupied. The attributes of an occupied voxel are the red, green, and blue components of the surface color. For each sequence, the cube is scaled so that it is the smallest bounding cube

that contains the entire sequence. Since the height of the subject is typically the longest dimension, for a subject 1.8 m high, a voxel at depth 10 would be approximately 1.8 m / 1024 voxels ≈ 1.75 mm per voxel on a side.

| *longdress* | *loot* | *redandblack* | *soldier* |

2. Availability

The terms of use of the dataset are governed by the License Agreement, which is an integral part of the dataset and which must accompany any copy of the dataset.

The dataset is initially made available in https://1drv.ms/f/s!AlI8AvELwyBbjMcLxPRgeVfuT6sFFQ as 8iVFBv2.zip. The location of the data may be moved to MPEG and/or JPEG servers. Zipped, the size is about 6 GB. Unzipped, the size is about 22 GB.

3. Folder Structure

The unzipped 8iVFBv2 folder contains the licensing agreement, this document, and one subfolder for each sequence:

- License.pdf and License.docx– 8i License Agreement for non-commercial use
- CiteThis.pdf and CiteThis.docx– This document, which should be cited in any publication
- longdress/
- loot/
- redandblack/
- soldier/

These items form the Dataset, and must be kept together.

The subfolder for each sequence (i.e., *seqname* in {*longdress*, *loot*, *redandblack*, *soldier*}) contains two subfolders and two regular files:

- Ply/ – main data, as separate ply file for each frame
  - *seqname*_vox10_MMMM.ply,…,seqname_vox10_NNNN.ply
- Png/ – images from orthogonal frontal projection for each frame
  - *seqname*_vox10_MMMM.png,…,seqname_vox10_NNNN.png
- *seqname*_vox10.mp4 – video constructed from png images
- Count.txt – text file containing number of voxels in each frame

The beginning frame # (MMMM) and ending frame # (NNNN) have the following values:

Sequence	Beginning frame #	Ending frame #	Number of frames
longdress	1051	1350	300
loot	1000	1299	300
redandblack	1450	1749	300
soldier	0536	0835	300

4. File Format

The main data are PLY files in ascii format [1], and if desired may be viewed directly in software such as MeshLab [2], as seen below.
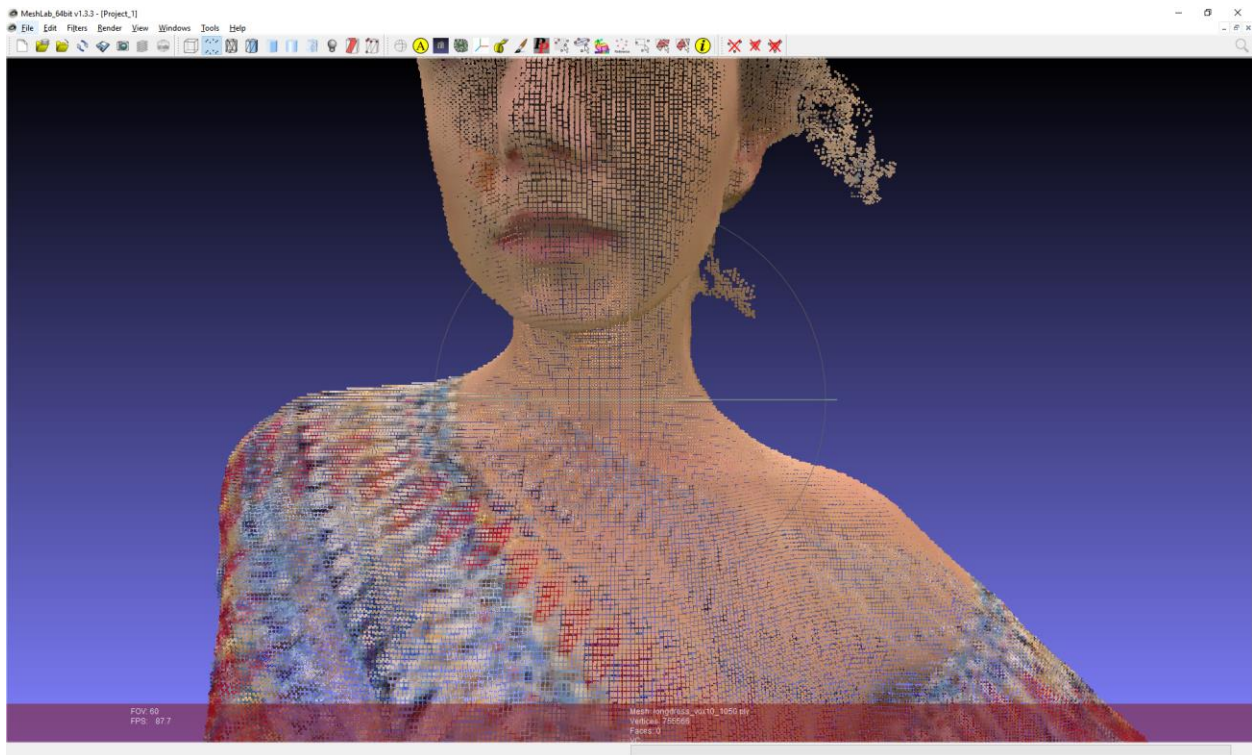
The beginning of longdress_vox10_1051.ply, for example, is:

```
ply
format ascii 1.0
comment Version 2, Copyright 2017, 8i Labs, Inc.
comment frame_to_world_scale 0.179523
comment frame_to_world_translation -45.2095 7.18301 -54.3561
comment width 1023
element vertex 765821
property float x
property float y
property float z
property uchar red
property uchar green
property uchar blue
end_header
211 63 61 133 104 77
210 63 63 128 97 69
211 62 63 126 97 71
211 63 62 130 100 73
211 63 63 129 98 71
212 63 59 139 110 83
213 63 59 140 112 85
214 63 58 147 120 94
214 63 59 140 113 87
215 62 59 138 111 84
215 63 58 147 121 94
215 63 59 141 114 87
213 61 63 122 95 71
212 63 60 136 107 80
212 63 61 133 104 77
.
.
.
```

Comment lines may be ignored by applications. The X, Y, and Z coordinates are integers in the range 0, …, 1023. 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) to the lower right rear of the subject from the subject's point of view, the side of the cube with the maximum X-coordinate to left of the subject, the side of the cube with the maximum Y-coordinate above the subject, and the side of the cube with the maximum Z-coordinate in front of the subject. That is, the X-axis points to the subject's left; the Y-axis points up, and the Z-axis points to the subject's front.

5. Processing

The data are produced by a capture system with 42 synchronized RGB cameras. The cameras are clustered into 14 groups, which are placed around the subject at a couple of meters' distance. Each group captures RGB as well as depth-from-stereo. The inputs from the clusters are fused into a

3D surface. For the purposes of this dataset, the surface is voxelized in a 1024x1024x1024 voxel cube. Each occupied voxel receives a representative color and is listed in the PLY file.

6. Citation

If you publish images or report performance results of these data, we request that you cite this document as Eugene d'Eon, Bob Harrison, Taos Myers, and Philip A. Chou, "8i Voxelized Full Bodies, version 2 – A Voxelized Point Cloud Dataset," ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document m40059/M74006, Geneva, January 2017.

7. References

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

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