Tampere University

Mikko Pitkänen

OPEN ACCESS DYNAMIC HUMAN POINT CLOUD DATASETS

Bachelor's Thesis
Faculty of Information Technology
and Communication Sciences
Examiner: Joonas Multanen
June 2023

ABSTRACT

Mikko Pitkänen: Open Access Dynamic Human Point Cloud Datasets
Bachelor's Thesis
Tampere University
Bachelor's Programme in Computing and Electrical Engineering
June 2023

Video conferencing tools in use today transmit 2-dimensional (2D) video. 2D video lacks the depth dimension, meaning the distance between an object and the camera. Therefore, 2D video may not be viewed from other angles in order to see behind captured objects. A more immersive form of tele-communication, known as tele-presence, instead utilizes volumetric video. Volumetric video is 3-dimensional (3D) video that can be viewed from any angle. Tele-presence systems create virtual spaces where the users are able to interact with the environment and with each other like in the real world.

Volumetric video may be implemented using point clouds. Point clouds are sets of sampled data points that represent surfaces in 3D space. However, point clouds require large amounts of storage. In order to encode point clouds into a less memory intensive format, point cloud compression is required. The development of tele-presence and point cloud compression technologies is currently lagging due to a lack of diverse test data.

This thesis evaluates the state of currently available point cloud datasets and their usability in development of the technologies mentioned above and compares the datasets. In order to ensure that the compared datasets match the planned use cases, a set of selection criteria is defined. In addition to the subjective visual quality of each dataset, the number of point clouds within each dataset and their contents are also considered.

As a result of the comparison, it is found that video produced using point clouds has lower visual quality than video produced using the commonly used textured meshes. The largest of the datasets, CWIPC-SXR, is found to be a versatile test dataset for tele-presence, due to the nature of the social interactions depicted in it. There still exists a very limited number of datasets matching the specified selection criteria and all except one contain only 1-4 point clouds. Therefore, more point cloud datasets meeting the criteria must be captured.

Keywords: Point Clouds, Volumetric Video, Multimedia

TIIVISTELMÄ

Mikko Pitkänen: Avoimet dynaamiset ihmispistepilvien datasetit
Kandidaatintutkielma
Tampereen yliopisto
Tieto- ja sähkötekniikan kandidaattiohjelma
Kesäkuu 2023

Nykyisin laajalti käytössä olevat videopuhelusovellukset välittävät äänen lisäksi myös kaksiulotteista (2D) videota. 2D-videosta puuttuu syvyysulottuvuus eli tieto kuvatun kappaleen ja kameran välisestä etäisyydestä. 2D-videota ei siis voi tarkastella toisesta kuvakulmasta niin, että kuvatun kappaleen taakse voisi nähdä. Sen sijaan etäläsnäolo (engl. tele-presence) hyödyntää volumetristä videota eli kolmiulotteista (3D) videota, jota voi tarkastella mistä tahansa kuvakulmasta. Etäläsnäolo on uusi todentuntuisempi telekommunikaation muoto, joka luo virtuaalisia tiloja, joissa käyttäjät näkevät itsensä ja toisensa virtuaalisissa kehoissa. Käyttäjät vaikuttavat toisiinsa ja virtuaaliseen ympäristöönsä, kuten oikeassakin maailmassa.

Volumetristä videota voidaan tuottaa käyttäen pistepilviä, jotka ovat yksi 3D-datan esitysmuodoista. Pistepilvet koostuvat joukosta toisiinsa kytkeytymättömiä pisteitä, jotka esittävät kappaleiden pintoja 3D-avaruudessa. Yksi pistepilvien suurimmista haasteista on niiden suuri muistinkulutus. Pistepilvet on pakattava vähemmän muistia kuluttavaan muotoon, jotta niitä voisi hyödyntää reaaliaikaisessa etäläsnäolon kaltaisessa sovelluksessa. Pistepilvien pakkausteknologioiden sekä uusien etäläsnäolojärjestelmien kehitys on 2D-videoon perustuvien sovellusten kehitystä hitaampaa kunnollisen testidatan puutteen vuoksi.

Tutkielman tarkoitus on tutkia saatavilla olevien datasettien tämänhetkistä tilaa ja soveltuvuutta edellä mainittujen teknologioiden kehittämisessä. Menetelmänä on kirjallisuuskatsaus, ja tutkielmassa vertaillaan avoimesti saatavilla olevia pistepilvien datasettejä. Vertailuun valittaville dataseteille määritellään kriteerit, jotta verrattavien datasettien sisällöt vastaavat mahdollisimman tarkasti suunniteltuja käyttötapauksia. Dataseteistä verrataan niiden sisällön lisäksi myös pistepilvien lukumääriä sekä subjektiivista kuvanlaatua.

Vertailun tuloksena todetaan, että pistepilvien avulla tuotettu video on kuvanlaadultaan heikompaa kuin yleisesti käytössä olevien teksturoitujen verkkojen (engl. textured meshes) avulla tuotettu video. Suurin dataseteistä, CWIPC-SXR, osoittautuu selvästi muita datasettejä sopivammaksi etäläsnäolojärjestelmien testaamisessa. Määriteltyjä kriteerejä vastaavia datasettejä on toistaiseksi olemassa vain vähän, ja yhtä lukuun ottamatta niissä oli vain 1–4 pistepilveä. Tämän perusteella kriteerit täyttäviä pistepilviä tarvitaan lisää.

Avainsanat: Pistepilvet, Volumetrinen video, Multimedia

PREFACE

This thesis was written during my stay as a research assistant in Ultra Video Group at Tampere University. I'd like to thank my instructor Joonas Multanen for feedback during the writing process. I'd also like to thank my friends and colleagues Guillaume Gautier, Alexandre Mercat, and Nyyti Saarimäki for their advice and guidance.

Tampere, 14 June 2023

Mikko Pitkänen

CONTENTS

LIST OF ABBREVIATIONS

2D	2-dimensional
3D	3-dimensional
8iVFBv2	8i Voxelized Full Bodies, version 2
8iVSLF	8i Voxelized Surface Light Field
G-PCC	Geometry-based point cloud compression
LIDAR	Light Detection and Ranging
MPEG	Moving Picture Expert Group
ODHMS	Owlii Dynamic Human Mesh Sequence
PCC	Point cloud compression
SXR	Social extended reality
V4	Volograms & V-SENSE Volumetric Video
V-PCC	Video-based point cloud compression

1. INTRODUCTION

Video conferencing tools have become increasingly important to many people in recent years, as they allow remote audiovisual communication. During the Covid-19 pandemic many activities shifted online, e.g., education and office work. Despite their active development, video conferencing tools in use today work by transmitting 2-dimensional (2D) video. However, as it has no depth dimension, it cannot be moved or rotated in order to reveal hidden parts of the captured scene. The next major development direction of video is investigating *volumetric video*, which contains 3-dimensional (3D) video frames. Volumetric video enables *tele-presence*: virtual spaces where all participants see each other in real time, even if they are physically far apart.

Volumetric video may be implemented with *point clouds*, which are representations of 3D data. Each point in a point cloud stores its own position, along with any other data that may be collected, such as surface color. The problem with point clouds is they contain large amounts of data. For example, using medium resolution point clouds for entertainment purposes requires over 3,6 Gbps of bandwidth. [1] Therefore, effective compression methods are needed.

Implementation of both tele-presence and *point cloud compression* (PCC) solutions is currently hindered by the lack of open access high-quality test data [2], [3]. Always testing with the same input data makes it simple to evaluate the correctness of the functionalities being developed. It is important that the test data features a broad range of real-world use cases. This helps to ensure that the right functionalities get tested.

In this thesis, I compare open access point cloud datasets, which could be used for tele-presence and video-based PCC tool development. In addition to the contents of the datasets, I will also examine the number of point clouds in each and their visual quality. The goal of this thesis is to evaluate the current state of point cloud datasets and identify areas for future development.

Section 2 describes the background and main motivation of the thesis. Section 3 presents different categories of point clouds and selects five datasets. This is followed by a comparison of the five point cloud datasets in Section 4. Lastly, Section 5 concludes this thesis.

2. TELE-PRESENCE

This section introduces tele-presence, the main motivator behind the comparison. It also introduces volumetric video, which may be used to implement tele-presence. Lastly, textured meshes and point clouds are introduced as possible ways to implement volumetric video.

Video conferencing tools in use today work by transmitting 2D video. The broadcaster of a video stream can move and rotate his own camera, influencing what is being captured and broadcasted. However, as 2D video has no depth dimension, it cannot be moved or rotated in order to reveal hidden parts of the captured scene. Therefore, the receivers of the video stream can only view the 2D video from the perspective it was originally captured from.

Tele-presence systems create virtual spaces where all participants see each other in real time, even if they are physically far apart. The participants are captured from multiple angles, allowing their shapes to be captured accurately. The 3D data representing the participants' bodies is segmented from the background and streamed to other participants. [4] Tele-presence systems allow the participants to interact with the environment, resulting in immersive experiences [1].

2.1 Volumetric Video

Traditional video is essentially a sequence of still 2D images called *frames*, where each frame represents the captured scene at one instance in time. Volumetric video, in turn, is a sequence of still 3D frames. [5] Volumetric video allows the viewer to move freely around the captured space and view it from different angles. Tele-presence may be implemented with volumetric video [1].

By utilizing multiple cameras and depth sensors, surfaces can be captured from multiple angles. 3D spatial coordinates can be extracted for each captured point using various techniques, such as stereo matching [4]. Using the resulting point positions, color data from the captured images can then be mapped to each 3D point, producing volumetric video [3].

2.2 Textured Meshes

Computer graphics pipelines traditionally render meshes as 3D objects. A mesh represents the surface of some object being rendered. Meshes are composed of many small polygons, usually triangles [6]. The renderer may also sample textures in order to apply images to the meshes.

Meshes may be used to implement volumetric video. Their advantages over point clouds include better visual quality, easier integration with computer graphics pipelines, and continuous surface representation. [7] However, meshes also need to store the connectivity information between the vertices of their polygons. Point clouds do not have such connections between points and may be read in any order.

2.3 Point Clouds

Point clouds represent surfaces of objects in 3D space [1]. A point cloud consists of a set of non-connected, or *discrete*, 3D data points. A point cloud contains only the sampled points, not every possible point that belongs on a surface. In addition to the spatial coordinates of each captured point, the points may contain any number of attributes. In some cases, the surface normal vector is also stored [8]. In order to render a point cloud as an image of the captured surface, the surface color must also be saved for each captured point. An example point cloud is depicted in Figure 1.

Figure 1. *Point cloud from [3] viewed from different positions.*

Point clouds are commonly used in mapping and land surveying to store the scanned model of the landscape. They are also used in autonomous vehicles in order to observe the environment surrounding the vehicle and avoid dangers [1]. Point clouds may also be used to implement volumetric video [2]. Their advantages over textured meshes include simpler data format and acquisition [6], [7].

2.3.1 Point Cloud Acquisition

Point clouds are produced by scanning, or *capturing*, 3D surfaces. In practice, the surfaces may be captured using LIDAR (Light Detection and Ranging) devices, depth sensors, and cameras. The number of points in a point cloud depends on the resolution of the capturing device and the rate at which its output is sampled. Point clouds may also be generated programmatically or captured inside computer simulations.

While capturing point clouds, the depth of the captured surface can be found using two types of methods:

1. **Passive methods** involve using multiple cameras to capture a surface from multiple angles. The cameras are calibrated and their positions relative to the surface

are known. Positions of points on the surface are then calculated using image matching in combination with the cameras' positions. Extra points may be generated by interpolating between pixels in the source images.

2. **Active methods** utilize more specialized hardware, where a light source sends a pulse of light toward a surface and then captures the reflected pulse on its way back. The surface's distance can be calculated based on the elapsed time between emitting and observing the pulse.

Passive and active methods may both be used to produce point clouds. They may also be used combined together. For example, *volumetric studios* produce high quality point clouds by augmenting point clouds obtained using passive methods with data obtained from active methods. [1]

2.3.2 Point Cloud Compression

In their raw form, point clouds consume large amounts of memory. For example, let's consider a point cloud that stores the spatial coordinates and surface colors of each captured point. Both attributes are stored as three floating point numbers (one for each of the point's 3 color components or 3D-coordinates). It follows that every point in the point cloud consists of six floats. If the resolution of the point cloud is such that one point represents an area of 1 mm^2 on the surface, a flat area of 1 m^2 would contain one million points. Therefore, on a platform where one float consists of two bytes, the point cloud described above would require 12 megabytes of memory. A tele-presence application capturing 30 point clouds per second would generate 360 megabytes of data each second. Clearly, some form of compression is needed.

The *Moving Picture Expert Group* (MPEG) is an international organization best known for its media coding standards. MPEG has previously defined many popular codecs, such as MP3 for audio and H.264 for video. In their meeting in 2013, MPEG discussed immersive tele-presence as a use case for point clouds. MPEG deemed point cloud compression necessary, due to the dynamic nature of points and the large amounts of data they generate, and began PCC standardization work. [1]

MPEG later published two PCC standards, *video-based PCC* (V-PCC) in 2020 and *geometry-based PCC* (G-PCC) in 2021. V-PCC is designed to encode video point clouds, which are sequences of individual point clouds. Each point cloud in the sequence represents the captured object at one point in time. G-PCC is designed to compress large individual point clouds with complex geometry. [9] As G-PCC deals with individual point

clouds and V-PCC with sequences of point clouds, the datasets being compared in this thesis are only suitable for V-PCC development.

2.3.3 Voxelized Point Clouds

Point clouds may be voxelized by dividing the 3D space into small cubical volumes called *voxels*. Each voxel takes its values from the points located within that volume. A voxel that contains no points is empty and does not belong in the voxelized point cloud. [1], [8], [10]

Voxelization speeds up processing by allowing voxels to be processed instead of single points. Voxelized point clouds also consume less memory, due to voxels getting their values from multiple sample points. The voxel grid can also be split into smaller grids to be processed in parallel.

3. DATASET SELECTION

This thesis compares open access point cloud datasets that are suitable for development and testing of tele-presence and V-PCC solutions. This section presents the criteria by which the datasets are chosen. It also briefly introduces the chosen datasets.

3.1 Selection Criteria

Point clouds may be categorized based on their density, content, and capture method:

- Density: Point clouds are categorized as either *sparse* or *dense point clouds*. LIDAR devices and depth sensors typically produce sparse point clouds. These point clouds are typically very precise, but their point density is rather low. [9] Sparse point clouds are used when low resolution is acceptable, such as mapping and surveying. Dense point clouds are typically produced using passive methods described in Section 2.3.1. Dense point clouds are often used when high resolution is desirable, such as in design and in digital conservation of antique artworks. [1]
- Content: Point clouds may be categorized as either *static* or *dynamic point clouds*. Static point clouds represent the captured surface at one point in time. Dynamic point clouds are sequences of static point clouds, which represent the surface over a period of time. Dynamic point clouds are generally used in virtual/augmented reality applications, such as tele-presence [1].
- Capture method: Point clouds are categorized as either *synthesized* or *natural point clouds*. Point clouds generated from 3D models or in computer simulations are synthesized point clouds. Point clouds captured in the real world using cameras and sensors are called natural point clouds.

Tele-presence, which motivates this comparison, is a real-world scenario which aims to enable immersive tele-communication experiences. In order to ensure the required level of quality and immersiveness, it should be tested with data similar to what it will eventually be used with. This means that natural point clouds should be preferred over synthesized point clouds. High immersion requires high levels of detail and visual quality, therefore dense point clouds should be favored over sparse point clouds. Tele-presence systems populate the virtual spaces with moving full body avatars, so the compared point

clouds should also be dynamic point clouds and represent whole human bodies. In summary, the compared datasets should all contain natural dense dynamic full-human point clouds.

3.2 Selected Datasets

Five datasets are selected for the comparison:

1. **CWIPC-SXR** is a dynamic human point cloud dataset released in [3]. It contains 21 different screenplays, which depict actors interacting in social extended reality (SXR) settings. CWIPC-SXR contains 45 point cloud sequences. Each sequence contains between 596 and 1384 frames and runs between 20 and 50 seconds. CWIPC-SXR is the largest of the selected datasets, partly because it also contains raw video files used to generate the point clouds. It is also the only dataset that was captured using consumer grade hardware, namely 7 Azure Kinect devices. Point cloud depicted in Figure 1 is from CWIPC-SXR.
2. **8i Voxelized Full Bodies, version 2** (8iVFBv2) is a dynamic human voxelized point cloud dataset released in [10]. It was contributed to MPEG as test material for future standardization efforts. 8iVFBv2 consists of 4 voxelized point cloud sequences, in which the actors perform small dance-like movements and strike poses. All of the sequences contain 300 frames and run for 10 seconds. The dataset was captured using 42 cameras arranged in 14 camera clusters around the subjects.
3. **8i Voxelized Surface Light Field** (8iVSLF) is another MPEG contribution released as potential test material in [8]. It contains a voxelized point cloud sequence, which features a female dancer. The sequence contains 300 frames and runs for 10 seconds. The dataset also contains 6 high-resolution static voxelized point clouds. 8iVSLF was captured using 39 cameras arranged in 13 camera clusters around the subjects. Instead of storing one color value per point, the point clouds in 8iVSLF store one color value per camera viewpoint.
4. **Owlii Dynamic Human Mesh Sequence** (ODHMS) is the third and final MPEG contribution in this comparison, released in [11]. It contains 4 mesh sequences. In the recorded sequences the actors exercise, play basketball, and dance. The sequences in ODHMS contain 600 frames each and run for 20 seconds. Owlii does not disclose which hardware was used to capture the dataset.

5. **Vologrames & V-SENSE Volumetric Video** (V4) is a dynamic human volumetric video dataset released in [2]. It contains 3 mesh sequences. In two of the sequences the actors perform quick dance moves and in the last one a long monologue. All 3 sequences in V4 contain between 150 and 1800 frames and run between 5 and 60 seconds. The dataset was captured using 12 cameras arranged around the subjects.

All selected datasets contain dynamic sequences featuring full bodies in point cloud or textured mesh formats. All sequences were captured at 30 frames per second.

4. DATASET COMPARISON

This section presents the comparison of selected datasets. It also analyses the results.

The selected datasets contain a total of 57 sequences, with the largest dataset containing 45 sequences. Most datasets only contain 3-4 sequences, and therefore satisfy rather narrow use cases only. Selected datasets and their parameters are summarized in Table 1.

Table 1. *Selected point cloud datasets and their parameters.*

Dataset	Format	Subjective visual quality	Sequences	Total actors	Total frames	Total filesize
CWIPC-SXR	Dynamic point cloud	Poor	45	23	44000	1,6 TB
8i Voxelized Full Bodies, version 2	Dynamic voxelized point cloud	Average	4	4	1200	22 GB
8i Voxelized Surface Light Field	Dynamic voxelized point cloud	Average	1	1	300	24 GB
Owlii Dynamic Human Textured Mesh Sequence	Dynamic mesh	Good	4	4	2400	40 GB
Volograms & V-SENSE Volumetric Video	Dynamic mesh	Good	3	3	2100	-

In subjective comparison, CWIPC-SXR has the poorest visual quality. The models have visible missing patches and overall look quite blocky. The 8iVFBv2 and 8iVSLF datasets look more realistic than CWIPC-SXR. The shapes are smoother and less blocky, but the texturing is slightly blurry. The ODHMS and V4 datasets have the best visual quality. The models look crisp and clear. In summary, among the compared datasets, meshes have better visual quality than point clouds.

CWIPC-SXR contains 21 different screenplays, 7 of which feature roles for multiple actors, with each actor's performance captured separately into its own point cloud sequence. In total, CWIPC-SXR contains 45 sequences and features 23 actors, 14 males and 9 females, dressed in a wide variety of outfits. 8iVFBv2 consists of 4 sequences. Each sequence features a different actor dressed in different colors and materials. Out of the 4 actors in 8iVFBv2, 2 are female and 2 are male. 8iVSLF contains a sequence, which features a female dancer in a dress. ODHMS consists of 4 sequences, all of which feature a different actor. One of the actors in ODHMS is female, and the rest are males. V4 contains 3 sequences, each of which has a unique actor. All sequences in V4 feature male actors with varying skin colors.

CWIPC-SXR features a wide range of social scenarios, e.g., man showing a book to the camera, man performing card tricks to another man, or 19 people dancing YMCA. In this regard, CWIPC-SXR is the most diverse dataset and well suited for development of tele-presence systems. Sequences in the rest of the datasets all feature a single actor performing solo actions, such as dancing or striking a pose. One exception to this is the sequence Sir Frederick in V4, which features an actor in medieval clothing giving a monologue lasting 60 seconds. The sequence could also be suitable for tele-presence development.

All point cloud datasets are suitable for V-PCC development. All of them feature sequences with actors doing fast movements in varying outfits. 8iVSLF presents details like facial expressions and hands accurately. The CWIPC-SXR dataset is also suitable for V-PCC development, but due to its lower original level of visual quality, artifacts introduced by faulty compression may not be as noticeable.

5. CONCLUSIONS

Future video conferencing solutions utilize volumetric video instead of traditional 2D video. Volumetric video may be implemented with point clouds, but they contain large amounts of data. In order to develop efficient point cloud compressors, high-quality test data is needed.

The goal of this thesis was to compare existing open access point cloud datasets and to evaluate the current state of point cloud datasets. The selected datasets were required to contain natural dense dynamic full-human point clouds. Through subjective comparisons, it was concluded that within the group of compared datasets, point cloud sequences have lower levels of visual quality compared to mesh sequences.

It was found that CWIPC-SXR is a versatile test dataset for tele-presence, due to the nature of the social interactions depicted in it. In addition, it was concluded that CWIPC-SXR is the least important test dataset for video-based point cloud compression, due to its poor visual quality.

There still exists a very limited number of natural dense dynamic full-human point cloud datasets and sequences. With the exception of one dataset, each dataset in the comparison contains only 1-4 sequences. In order to generate more test data to facilitate tele-presence development, more social interactions should be captured in high-quality point clouds.

REFERENCES

[1] D. Graziosi, O. Nakagami, S. Kuma, A. Zaghetto, T. Suzuki, and A. Tabatabai, “An overview of ongoing point cloud compression standardization activities: video-based (V-PCC) and geometry-based (G-PCC),” *APSIPA Trans. Signal Inf. Process.*, vol. 9, no. 1, 2020, doi: 10.1017/ATSIP.2020.12.

[2] R. Pagés, K. Amplianitis, J. Ondrej, E. Zerman, and A. Smolic, “Volograms & V-SENSE Volumetric Video Dataset,” 2022, doi: 10.13140/RG.2.2.24235.31529/1.

[3] I. Reimat, E. Alexiou, J. Jansen, I. Viola, S. Subramanyam, and P. Cesar, “CWIPC-SXR: Point Cloud dynamic human dataset for Social XR,” in *Proceedings of the 12th ACM Multimedia Systems Conference*, Istanbul Turkey: ACM, Jun. 2021, pp. 300–306. doi: 10.1145/3458305.3478452.

[4] S. Orts-Escolano, C. Rhemann, S. Fanello, W. Chang, A. Kowdle, Y. Degtyarev, *et al.*, “Holoportation: Virtual 3D Teleportation in Real-time,” in *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, Tokyo Japan: ACM, Oct. 2016, pp. 741–754. doi: 10.1145/2984511.2984517.

[5] L. Ilola, L. Kondrad, S. Schwarz, and A. Hamza, “An Overview of the MPEG Standard for Storage and Transport of Visual Volumetric Video-Based Coding,” *Front. Signal Process.*, vol. 2, p. 883943, Apr. 2022, doi: 10.3389/frsip.2022.883943.

[6] S. Perry, A. Pinheiro, E. Dumic, and L. A. Da Silva Cruz, “Study of Subjective and Objective Quality Evaluation of 3D Point Cloud Data by the JPEG Committee,” *Electron. Imaging*, vol. 31, no. 10, pp. 312-1-312–7, Jan. 2019, doi: 10.2352/ISSN.2470-1173.2019.10.IQSP-312.

[7] E. Zerman, C. Ozcinar, P. Gao, and A. Smolic, “Textured Mesh vs Coloured Point Cloud: A Subjective Study for Volumetric Video Compression,” in *2020 Twelfth International Conference on Quality of Multimedia Experience (QoMEX)*, May 2020, pp. 1–6. doi: 10.1109/QoMEX48832.2020.9123137.

[8] M. Krivokuća, P. A. Chou, and P. Savill, “8i Voxelized Surface Light Field (8iVSLF) Dataset,” presented at the ISO/IEC JTC1/SC29 WG11 (MPEG) input document m42914, Ljubljana, Jul. 2018.

[9] C. Cao, M. Preda, V. Zakharchenko, E. S. Jang, and T. Zaharia, “Compression of Sparse and Dense Dynamic Point Clouds—Methods and Standards,” *Proc. IEEE*, vol. 109, no. 9, pp. 1537–1558, Sep. 2021, doi: 10.1109/JPROC.2021.3085957.

[10] E. d’Eon, B. Harrison, T. Myers, and P. A. Chou, “8i Voxelized Full Bodies, version 2 – A Voxelized Point Cloud Dataset,” presented at the ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document m40059/M74006, Geneva, Jan. 2017.

[11] Y. Xu, Y. Lu, and Z. Wen, “Owlii Dynamic Human Textured Mesh Sequence Dataset,” presented at the ISO/IEC JTC1/SC29/WG11 m41658, 120th MPEG Meeting, Macau, Oct. 2017.