



# Volumetric Video配信での 視聴者仮想空間位置情報を 応用したデータ削減の一検討

NTT ネットワーク基盤研究所 趙笑添, 奥山隆文

## ○ Volumetric Videoとは

- Volumetric Videoの紹介とユースケース
- Volumetric Video制作技術
- Volumetric Videoデータと特性
- Volumetric Videoの再生技術

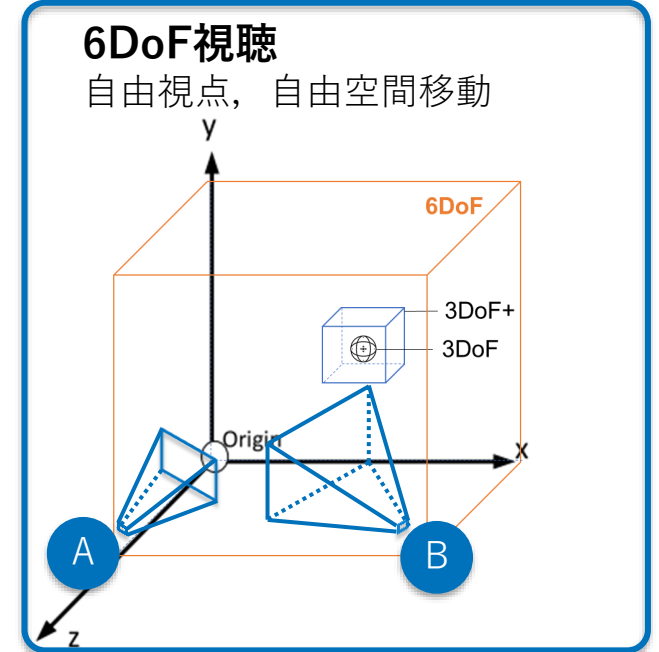
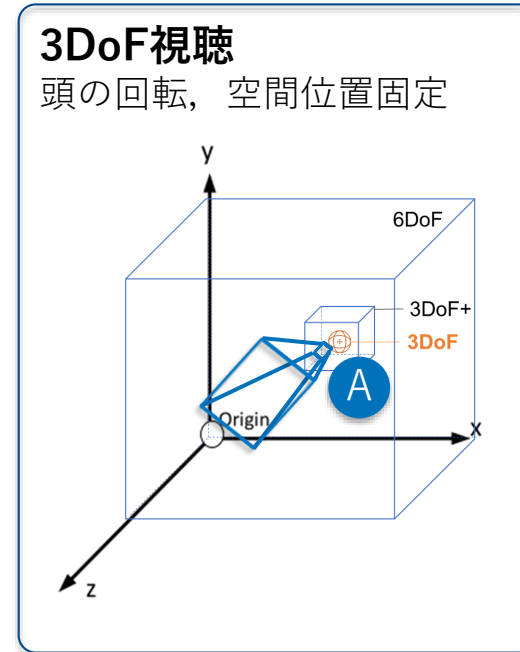
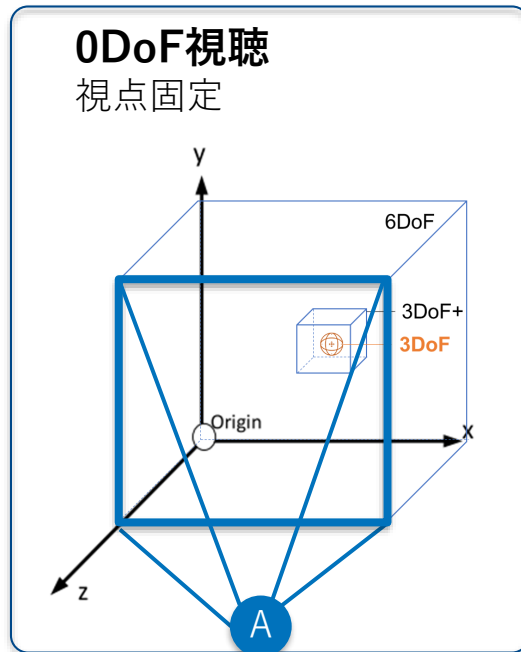
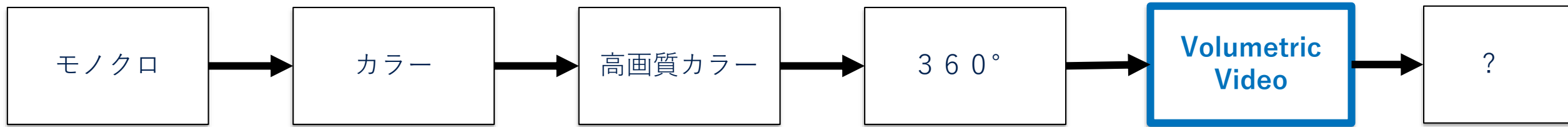
## ○ Volumetric Video配信へ <NWプラットフォームとして>

- 提案：視聴者情報を用いたデータ削減手法
- 提案手法定量評価
- まとめ

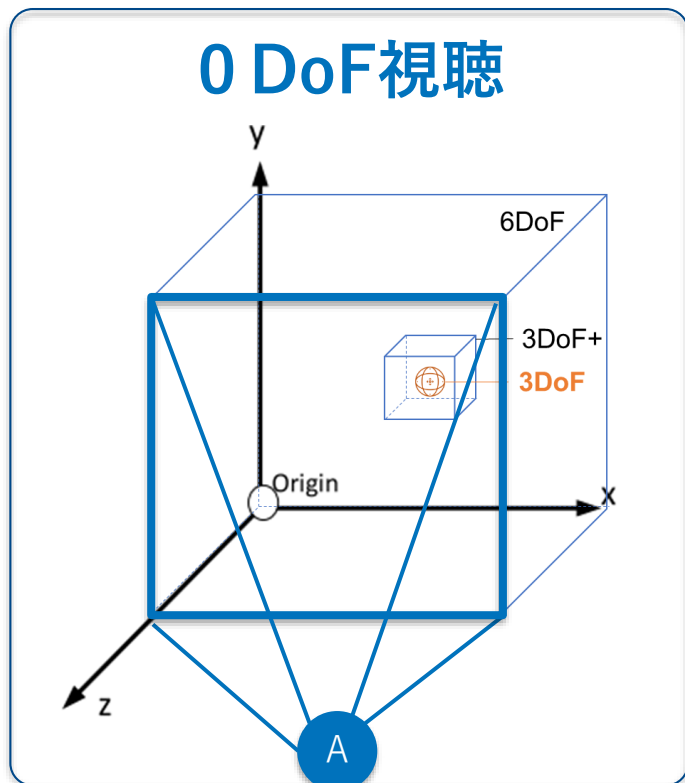
# Volumetric Videoとは

# より臨場感のある映像体験へ

- 臨場感の高いメディアへの追求
- Volumetric Videoは新しい高臨場コンテンツ
- 空間の中に入り自由視点での視聴体験が可能 (6DoF)

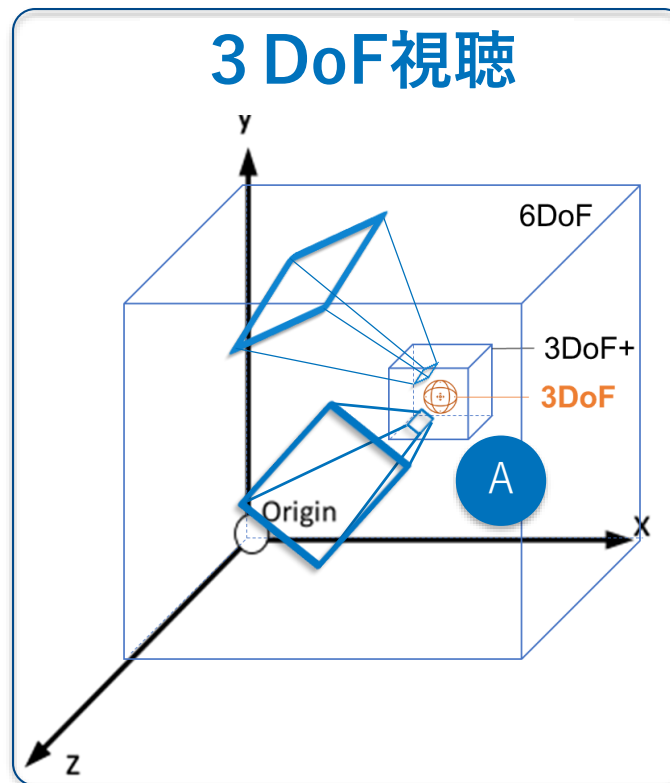


# より臨場感のある映像体験へ



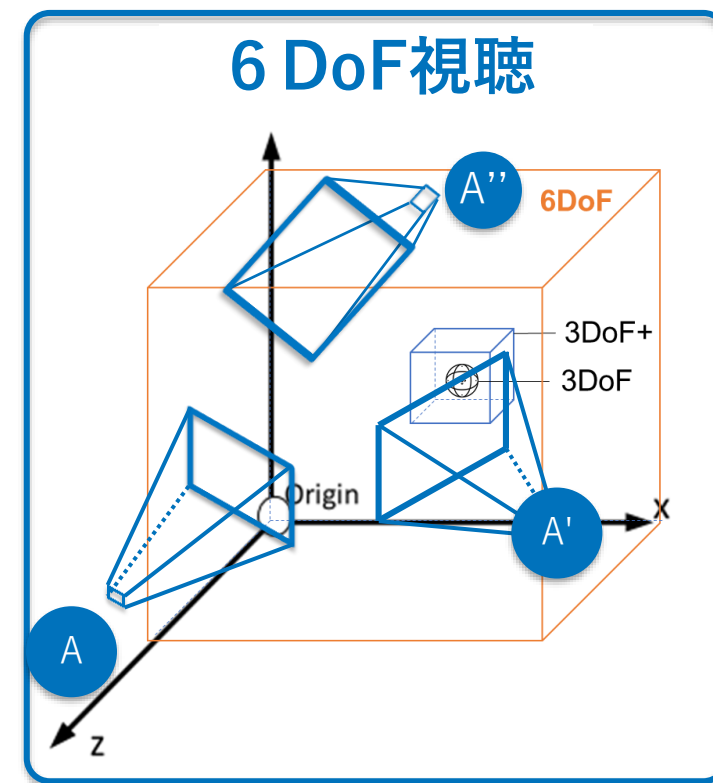
## 固定視界

固定地点・固定視界  
での視聴



## 360° 自由視界

固定地点から頭を動かし、  
360度自由視界での視聴



## 完全自由視界

自由視界，自由空間移動  
移動と視界に制限がない

# Volumetric Video で可能になること



配信V.V.映像



好きなCG環境



CG空間の中でV.V.映像を見る。  
動き回る。



Volumetric Video再生例

Data from <https://www.crescentinc.co.jp/>

# Volumetric Videoの利用シーン

Volumetric Videoは、6自由度での高臨場体験

既存のエンターテインメントだけでなく、  
将来の人と人の交流手段の一つになりうる



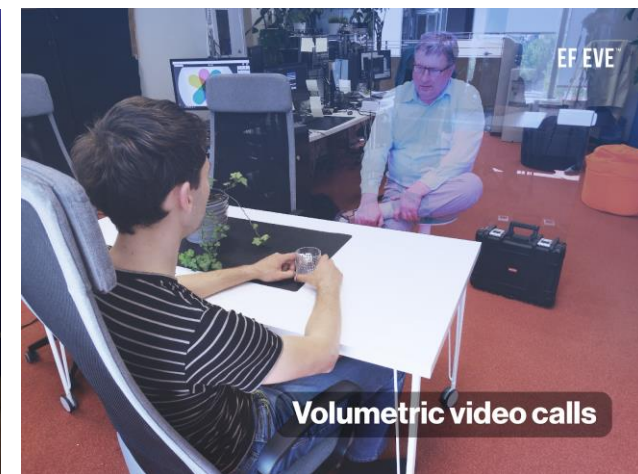
スポーツ



ドラマ



コンサート



会議・チャット



# Volumetric Videoの市場動向

各業界がVolumetric Videoに注目

- 通信キャリア業界(AT&T, Verizon等)
- クラウドサービス業界(Google, Microsoft, Facebook等)
- デバイス業界(Intel, Sony, HTC, MagicLeap等)

市場規模は14億ドルであり、2025年には4倍以上の58億ドルになると見込まれる

今後Volumetric Video は広く使われるコンテンツに成長する可能性

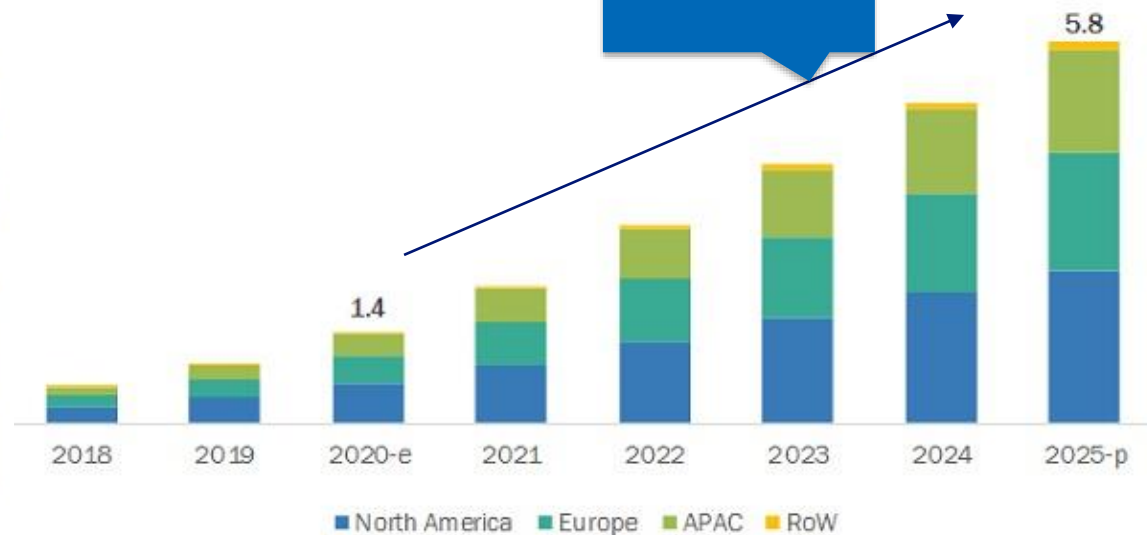
2025年に  
4倍増加



AT&T Shape の展示  
(AT&T)



Holoportation  
(Microsoft)



市場規模の地域ごとの変化予測

# NTTの取り組み

## NTTコミュニケーションズ

- 仮想現実ソリューションを展開
- NTT Communications Forum 2018  
Volumetric Videoの演劇、音楽ライブ、アトラクション分野への応用提案

## NTTドコモ

- XRを使ったサービス開始
- XRデバイスでのインタラクティブ展示
- Volumetric Video撮影スタジオの提供
- Docomo Open House2021  
Volumetric Videoストリーミング配信を展示



Volumetric Videoでの狂言



ドコモ×国立科学博物館  
XRで楽しむ未来の展示



ドコモXRスタジオ  
(Volumetric Video 撮影スタジオ)



VRバドミントン  
(Volumetric Video)

# Volumetric Video撮影から再生まで



Volumetric Videoの撮影から

AR/VRデバイスで再生には、具体的に5ステップある。

撮影

V.V.データ生成

V.V.データ  
保存・読み込み

3D  
レンダリング

Display表示

# V.V.の撮影

撮影

V.V.データ生成

V.V.データ  
保存・読み込み

3D  
レンダリング

Display表示



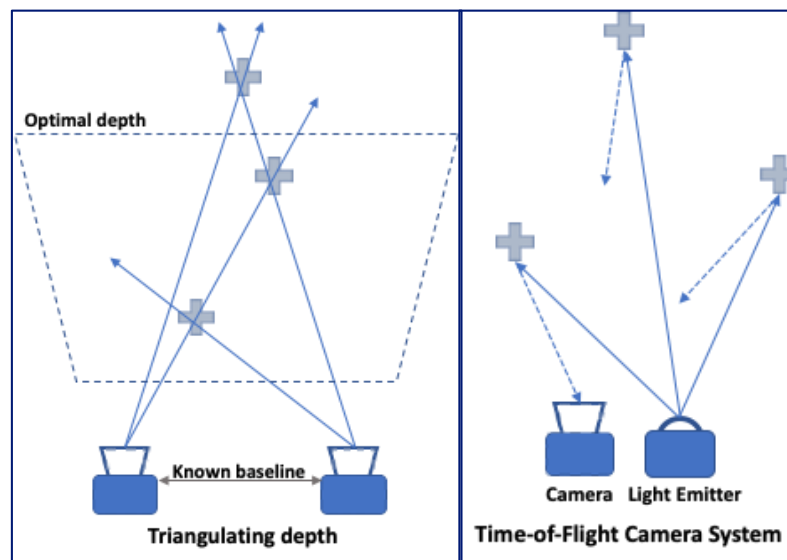
Volumetric Videoの撮影は、被写体を取り囲む多方向のカメラを配置することが必要。  
撮影のカメラはRGBカメラ、Depthカメラの組み合わせで制作される。

Volumetric VideoのDepthカメラは、以下の手法がある。

- **Stereo** : Intel RealSense, Stereolab ZED2など  
両眼カメラを用い、右左目のピクセルの画素位置の差から物体距離を算出
- **Time of Flight (ToF)** : Microsoft Azure Kinect, LiDARなど  
LED,IRやLaser光を照射し、物体で反射し戻ってくるまでの時間を計測することで距離を算出

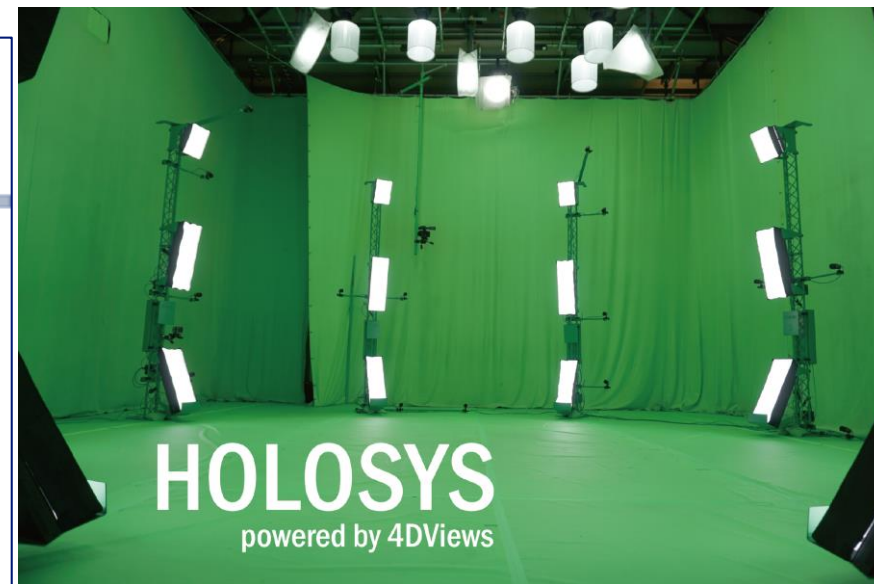


RGB-D例 :  
Microsoft Azure Kinect



ステレオカメラ

ToF方式カメラ



大規模V.V. 全方位撮影スタジオ例

# V.V.のデータ生成

撮影

V.V.データ生成

V.V.データ  
保存・読み込み

3D  
レンダリング

Display表示



RGB-Dセンサーでの撮影後、以下のパイプラインで3D点群および3Dメッシュが生成

撮影

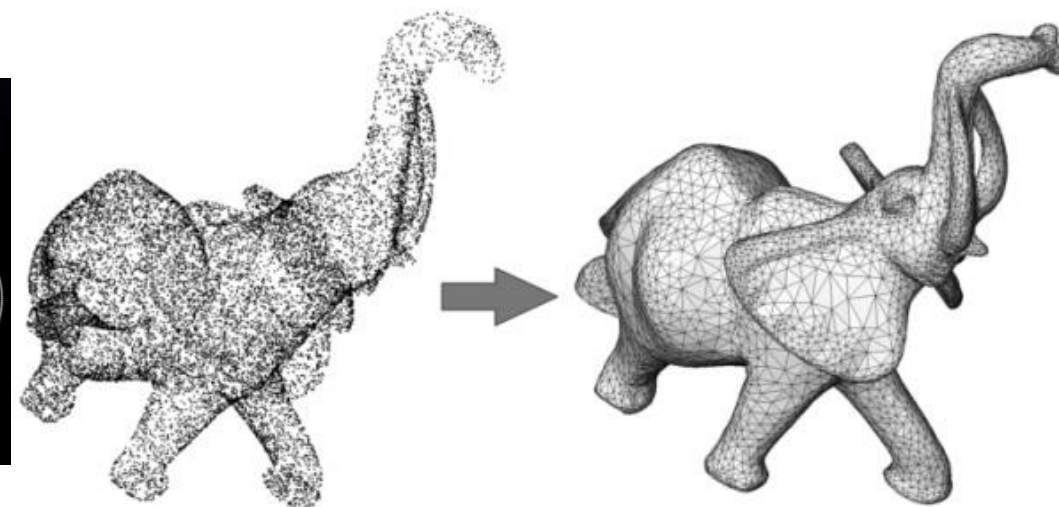
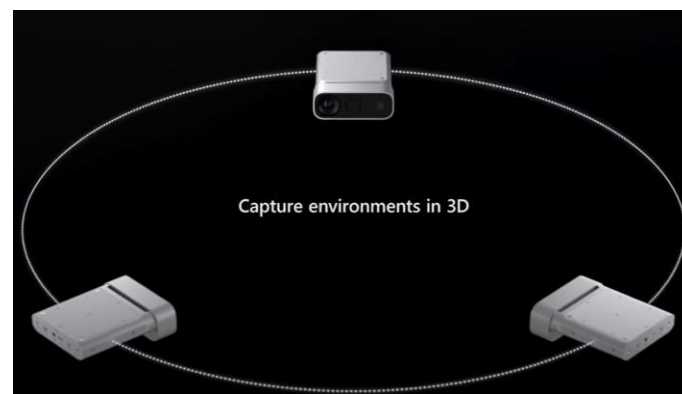
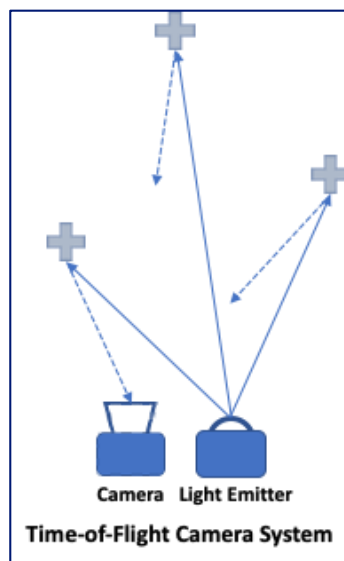
深度  
推定

複数カメラ  
深度データ合成

3D  
点群

サーフェス  
再構築

3D  
メッシュ



ポアソンサーフェス再構築手法

# V.V.制作スタジオ

撮影

V.V.データ生成

V.V.データ  
保存・読み込み

3D  
レンダリング

Display表示



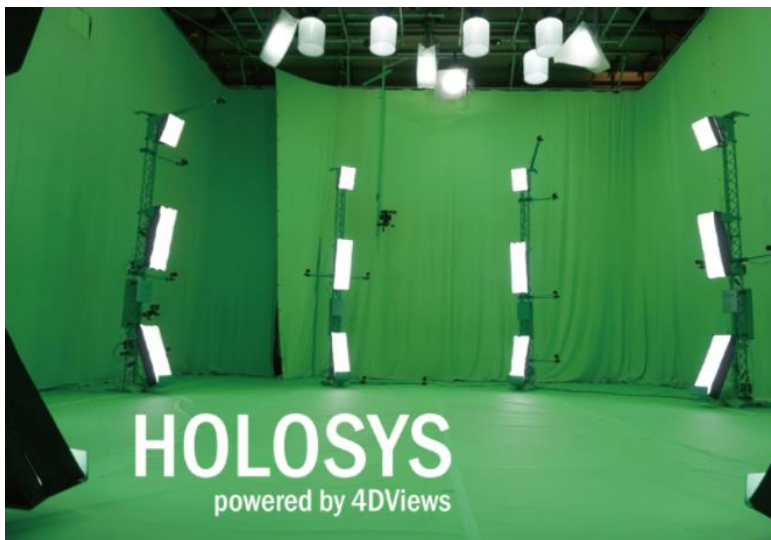
既存のコンテンツ制作スタジオソリューションは、

## 商用大規模スタジオ

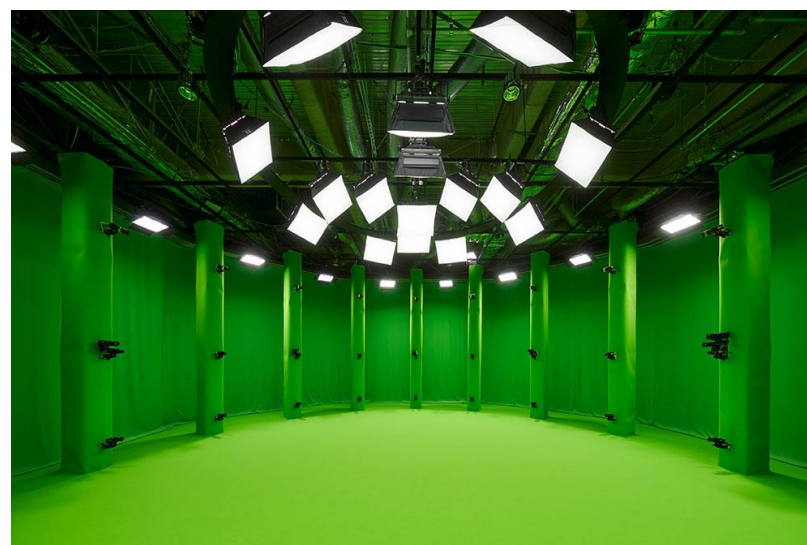
- 4D Views HOLOSYS
- 8i ASPX
- Microsoft Mixed Reality Capture Studio
- Sony<非公開>
- Canon <非公開>

## DIYセットアップスタジオ

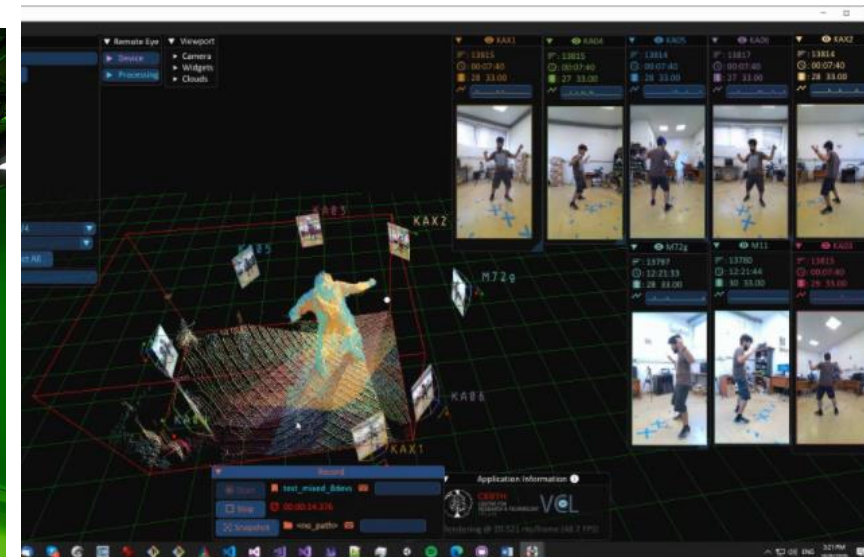
- EF EVE (商用)
- VCL3D/VolumetricCapture (Open Source)



4DViews システム



SONY システム



VCL3D/VolumetricCapture システム

# V.V.の3Dデータ

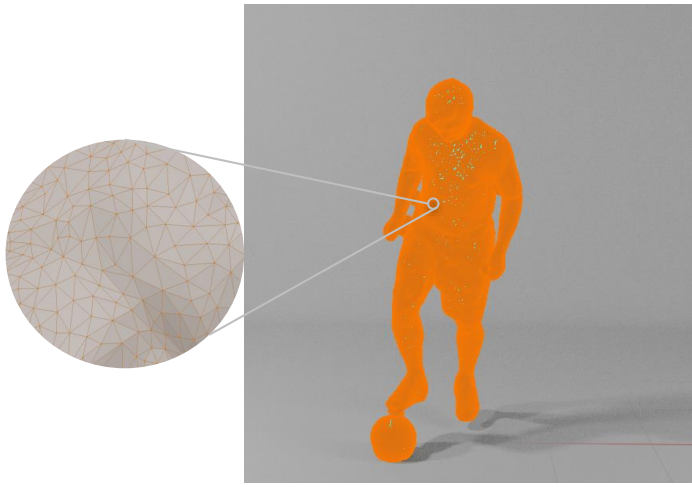
撮影

V.V.データ生成

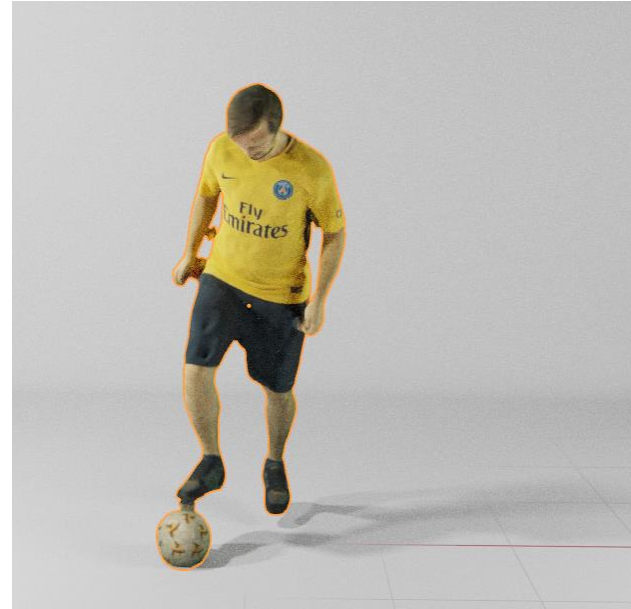
V.V.データ  
保存・読み込み

3D  
レンダリング

Display表示



3Dメッシュデータ



テクスチャーデータ

- **Wavefront Obj (.obj)**  
text/plain形式, 1frame/file  
CG映画の3Dアニメーション向けに開発
- **GL Transmission Format (.gltf, .glb)**  
jsonまたはbin形式, 1frame/file  
Webのやり取り、表示を前提に開発
- **Alembic (.abc)**  
bin形式, 複数frames/file  
CG映画の3Dアニメーション向けに開発

## 画像形式

- Portable Network Graphics(.png)
- JPEG

## 動画形式

- MP4

# V.V. 3Dメッシュデータ量

撮影

V.V.データ生成

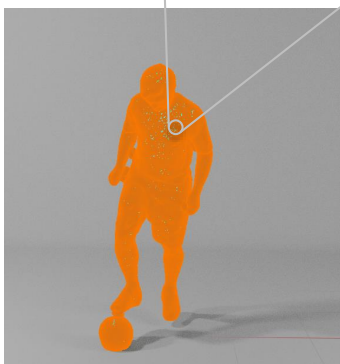
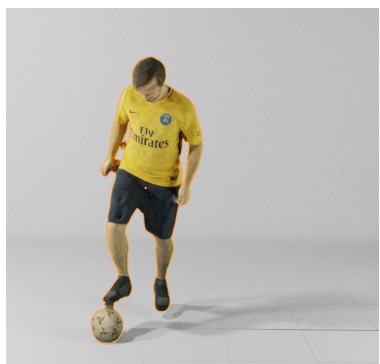
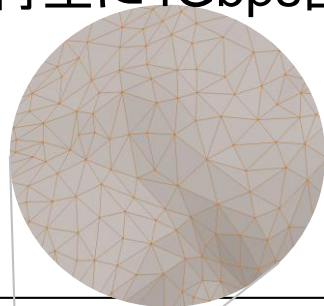
V.V.データ  
保存・読み込み

3D  
レンダリング

Display表示

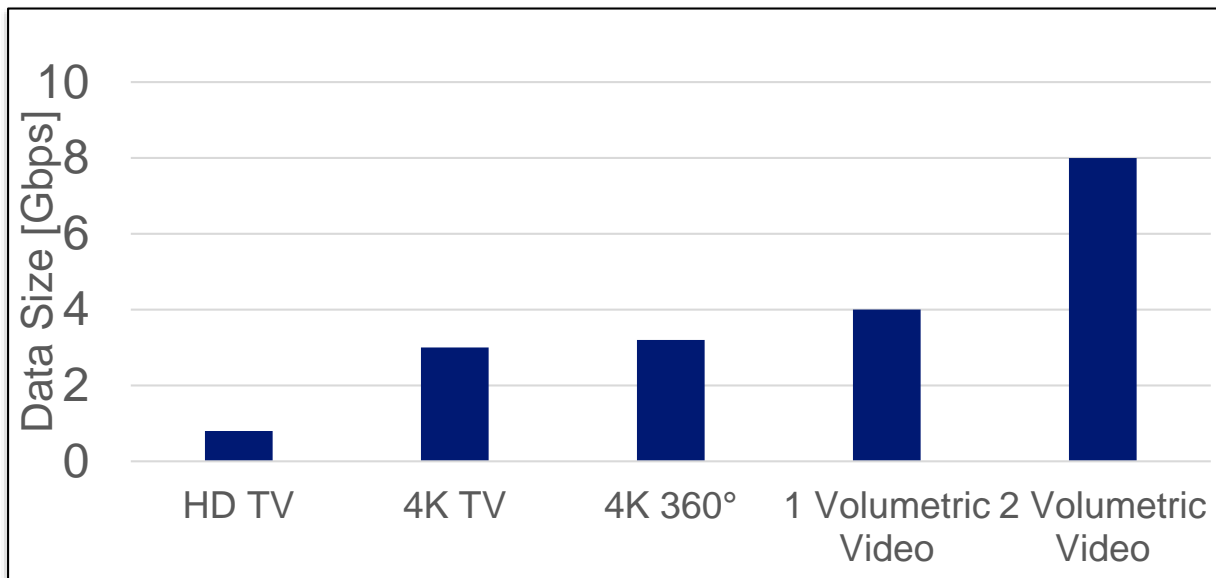


- Time Varying Mesh : 1フレームごとに新しいデータ
- 1 Volumetric Video再生に4Gbps回線



V.V. Data = Mesh + Texture  
4 MB/frame      3.1 MB      0.9 MB

1フレームのV.V.3Dメッシュデータ量(.obj + .jpg)



非圧縮時のデータ量(2D動画はInterlaceありの場合)[2]



# 3Dレンダリング

撮影

V.V.データ生成

V.V.データ  
保存・読み込み

3D  
レンダリング

Display表示



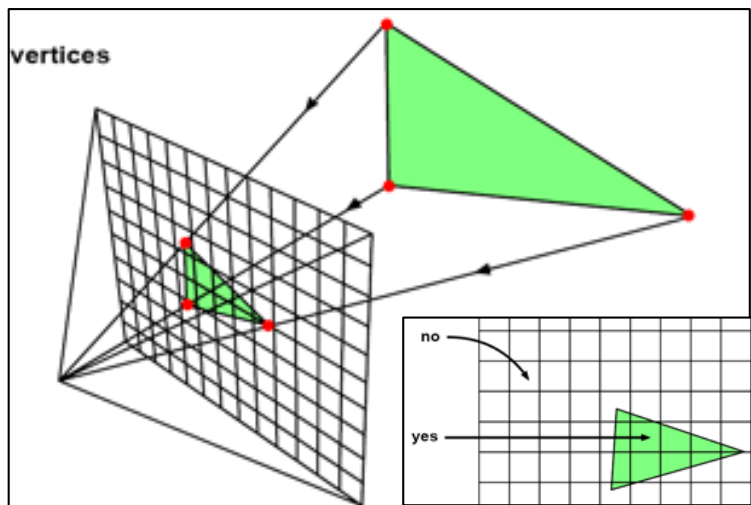
伝送された3Dモデルを他の3Dオブジェクトと合わせて、ディスプレイで表示するために2Dに変換をするレンダリングで主に使われる手法は、

## ○ Rasterization

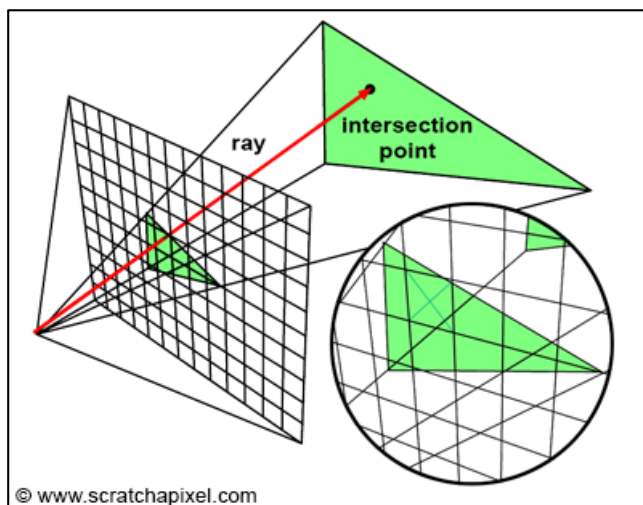
視界から見える3Dオブジェクトの色情報をそのまま2Dで反映  
処理コストが少ない、リアルタイムで生成できる。  
利用先：リアルタイム処理のゲームなどが多い。

## ○ Ray tracing

視界に届く光線を逆にたどることで、光の反射、屈折、光沢の表現ができる。  
処理コストと時間は、大幅に多いが、現実と区別がつかないほどの写実感を実現。  
利用先：映画、CG画像、（リアルタイムゲームでも技術的に可能となりつつある）



Rasterization Method



Ray Tracing



IKEAのカatalog画像（CG制作）

# 再生

撮影

V.V.データ生成

V.V.データ  
保存・読み込み

3D  
レンダリング

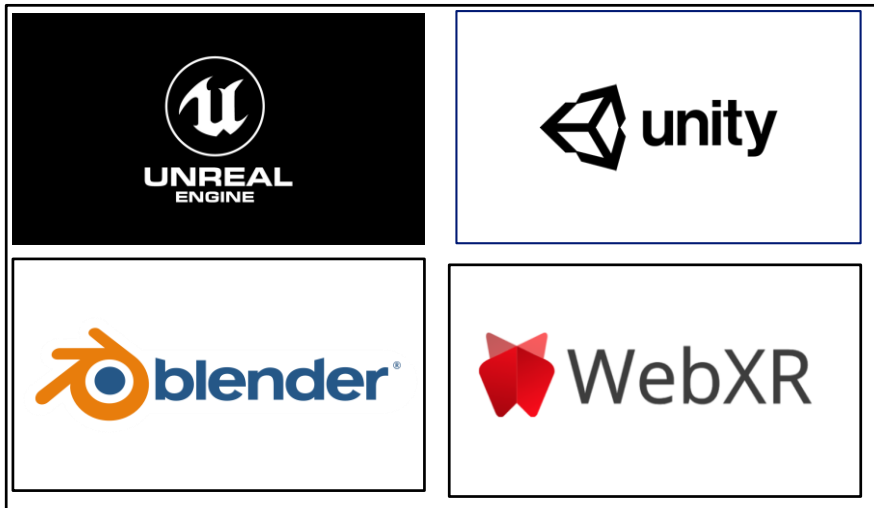
Display表示



Volumetric Videoおよび3Dメッシュデータの再生には、代表的に以下のツールがある。  
これらが、VR・MRデバイスから空間位置、頭の動きを受け取り、3Dレンダリングを行い、2D映像をデバイスに伝送する。

- **Unreal Engine**  
商用ゲーム開発ソフト
- **Unity**  
商用ゲーム開発ソフト

- **Blender**  
OpenSource  
3D編集ソフト
- **WebXR**  
OpenSource  
Webブラウザ上でのVRコンテンツの基盤API



レンダリング・ゲームエンジン  
Copyright 2020 NTT CORPORATION

空間位置、  
頭の動き  
←

2D視界映像  
→



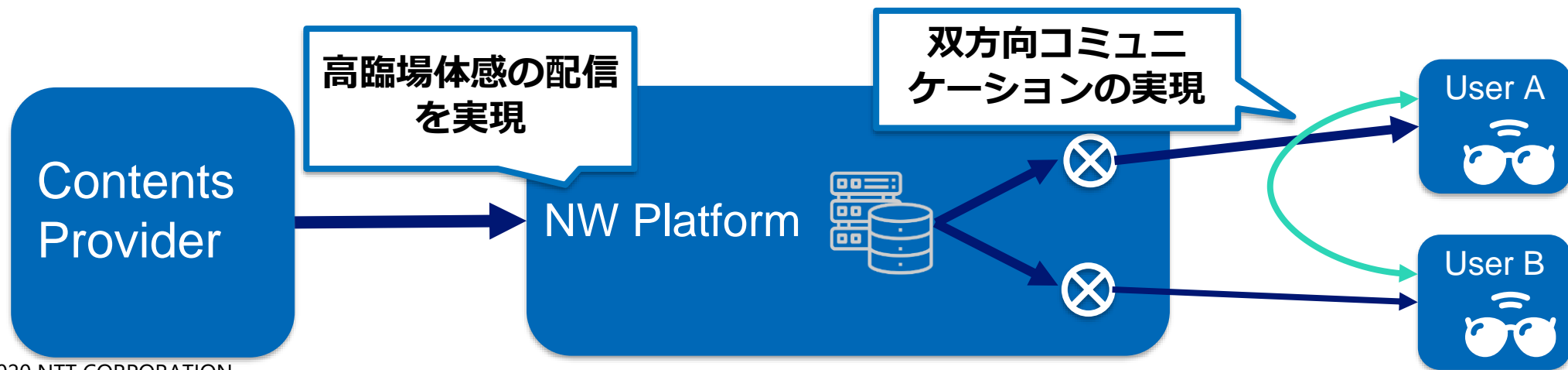
[https://www.nttdocomo.co.jp/product/magicleap1/images/thumb\\_magicleap1.png](https://www.nttdocomo.co.jp/product/magicleap1/images/thumb_magicleap1.png),  
[https://images-na.ssl-images-amazon.com/images/I/61HhQOJILOL.AC\\_SX522.jpg](https://images-na.ssl-images-amazon.com/images/I/61HhQOJILOL.AC_SX522.jpg),  
<https://www.playstation.com/ja-jp/ps-vr/>,  
[https://images-na.ssl-images-amazon.com/images/I/51qWjFuAShL.AC\\_SX425.jpg](https://images-na.ssl-images-amazon.com/images/I/51qWjFuAShL.AC_SX425.jpg)

# Volumetric Videoの配信へ NWプラットフォームとして

# ゴール | 配信NWプラットフォームの実現

- コンテンツプロバイダー  
⇒当配信NWプラットフォームの利用で、配信システム構築コストを削減
- 配信NWプラットフォーム  
⇒視聴者の高臨場体感を実現  
⇒視聴者同士が配信コンテンツを見ながら、コミュニケーション可能な双方向通信を実現
- 視聴者  
⇒FTTH・5Gのアクセス回線(100~1000Mbps)でもVolumetric Videoの高臨場視聴  
⇒他視聴者とVolumetric Videoでコミュニケーション

配信NWプラットフォームとして配信することで効率的配信を実現



# V.V.配信プラットフォーム

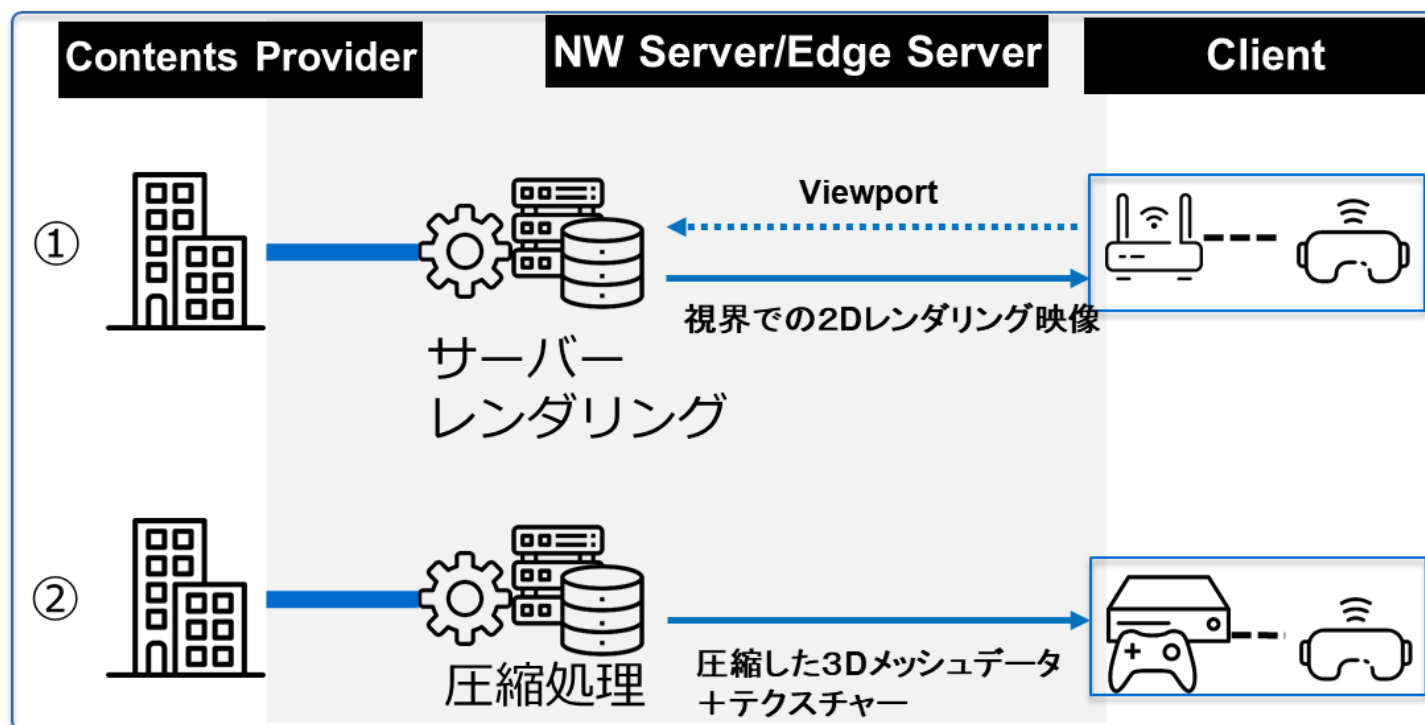
Volumetric Videoの配信手法には、

## ① サーバーレンダリング方式

クライアントから視界情報を受け取り、リアルタイムにサーバーでレンダリング生成レンダリング2D映像をクライアントに伝送

## ② クライアントレンダリング方式

3Dメッシュデータをクライアントに伝送、クライアントのPCでレンダリングそのままではデータが大きいため、圧縮後伝送されることが多い



# 既存のV.V.配信プラットフォーム

既存のV.V.配信は、

## ① サーバレンダリング方式

毎ユーザーごとにレンダリング，高い遅延要件のものには向かない (VRの全体処理許容遅延20ms達成が難しい)，サーバ負担大  
⇒ **同時使用者数とNWの帯域変動が視界映像全体に影響**

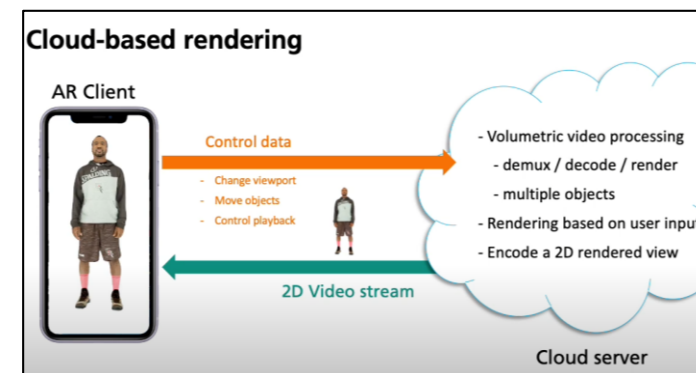
## ② クライアントレンダリング方式

**3Dデータ直接配信** [Orts-E. et al. (Holoportation)]

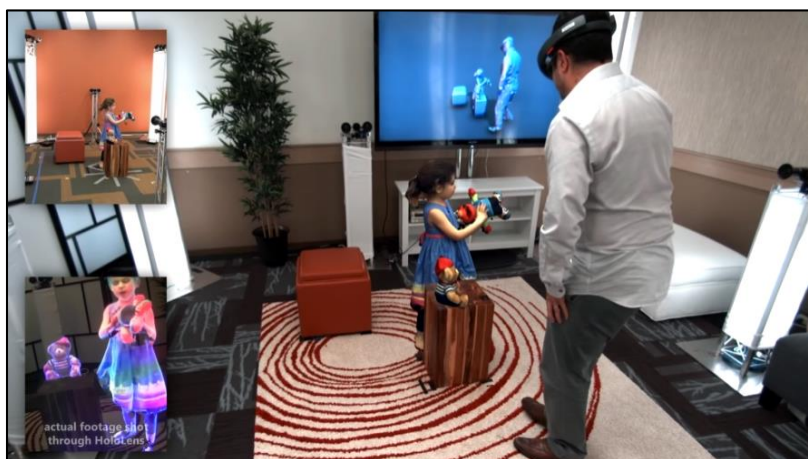
⇒再生に4Gbps必要，消費者向けの配信での適用が難しい

**3Dデータ全体のLevel Of Detailの帯域に応じたデータ削減**[Holostream]

⇒コンテンツ全体の品質低下により体感品質に影響



Gül, S., et al (Fraunhofer HHI)



[Gül, S., et al] Gül, S., et al, (2020). Low-latency cloud-based volumetric video streaming using head motion prediction. *NOSSDAV 2020 - Proceedings of the 2020 Workshop on Network and Operating System Support for Digital Audio and Video*

[Orts-E. et al. ] Orts-Escolano, et al. (2016). Holoportation: Virtual 3D teleportation in real-time. *UIST 2016 - Proceedings of the 29th Annual Symposium on User Interface Software and Technology, UIST '16*, 741–754

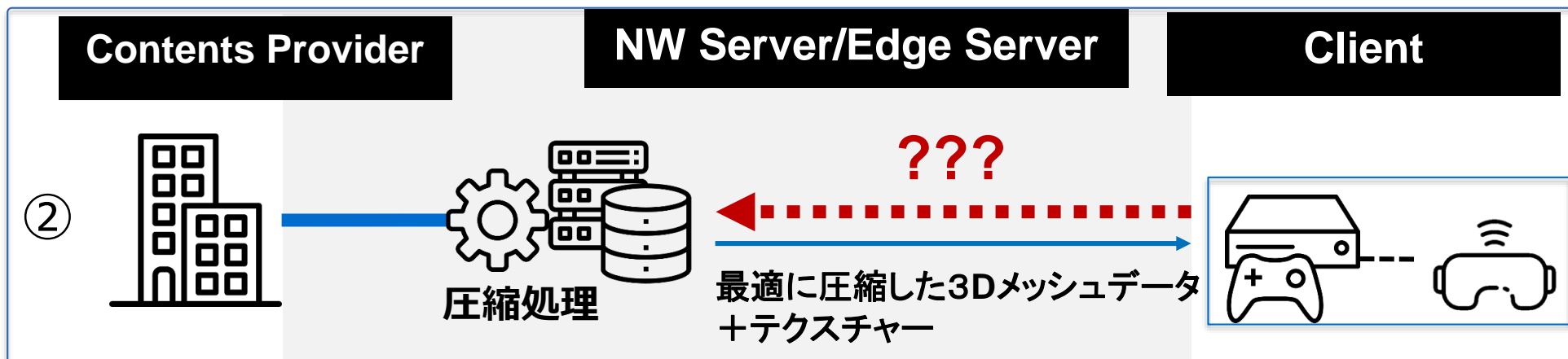
[Holostream] <https://arcturus.studio/holostream>

# 既存のV.V.配信プラットフォームでの課題 NTT

既存のクライアント配信手法では、  
圧縮を用いているが、ユーザーからの情報をほとんど使われていない。



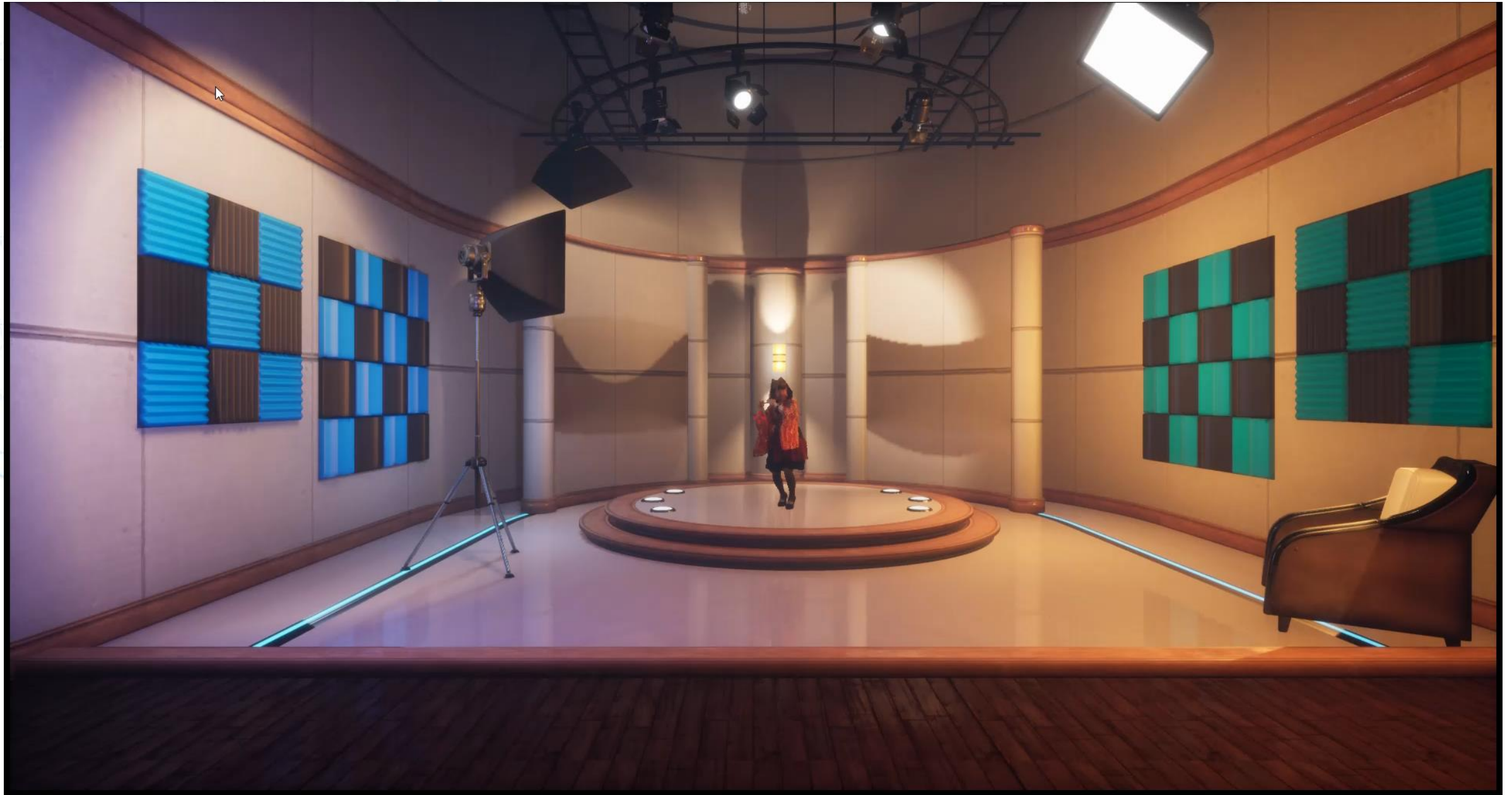
ユーザーからの情報をより有効的に活用したデータ圧縮手法  
があるのではないか？



提案：

視聴者仮想空間位置情報による  
データ削減手法





Volumetric Video再生例

Data from <https://www.crescentinc.co.jp/>

3Dデータは、  
3Dでないといけなののか？

# 提案手法の概要

- 視聴時, 3Dデータは全面3Dである必要はない  
再生時, 常に一面のみの3D映像が見え, 背面は視聴者から見えない
- Volumetric Videoの再生データの背面部は視聴に影響なく削減できる可能性  
⇒4ステップの提案手法
- 視聴者の空間位置情報から、リアルタイムにサーバーで配信一面3Dデータを作成
- 一面3Dデータであるため、通常の半分のデータ量で送信が可能

3D空間  
位置取得

V.V.視聴範囲の  
選択

リアルタイム  
一面3Dデータ  
生成

データ圧縮

# 3D空間位置取得・V.V.視聴範囲の選択

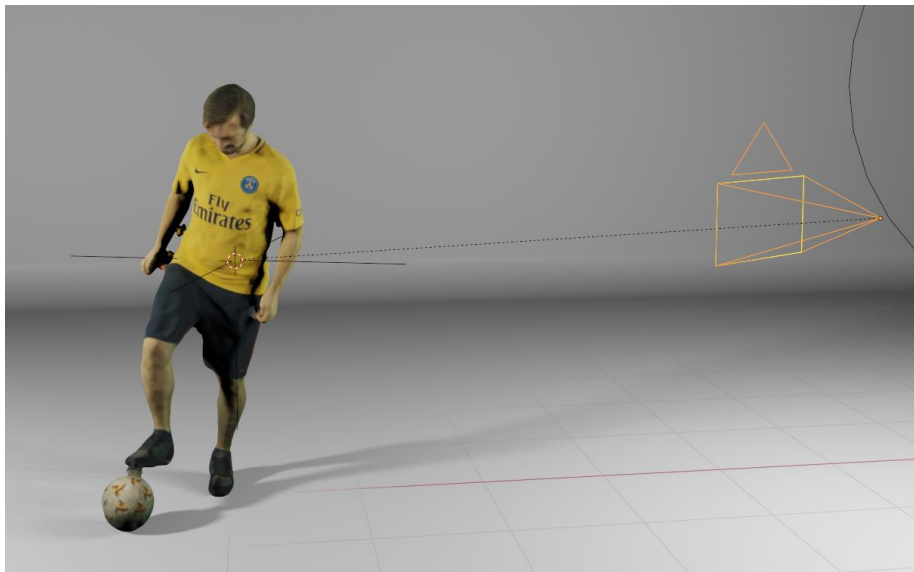
3D空間  
位置取得

V.V.視聴範囲の  
選択

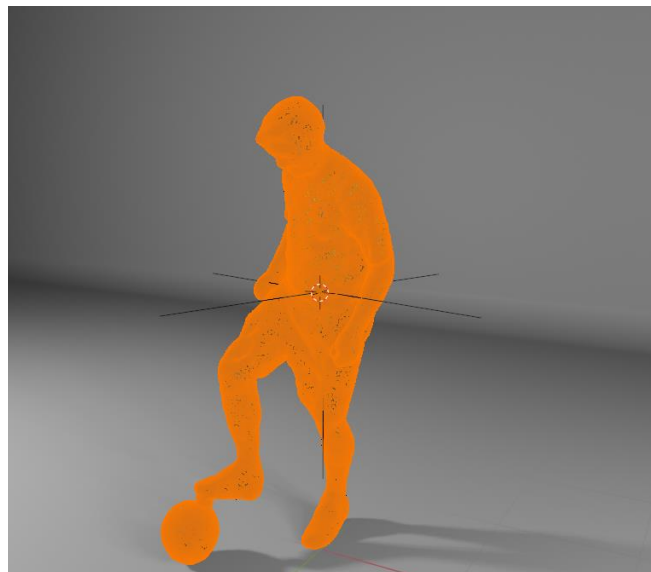
リアルタイム  
一面3Dデータ  
生成

データ圧縮

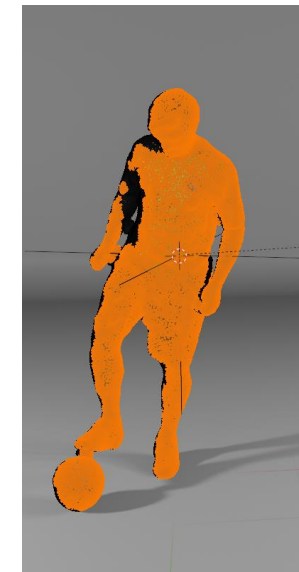
- VRデバイスから、視聴者の仮想空間上での位置と頭の向きを取得
- 視界にある3Dメッシュデータの内、視聴範囲内の頂点を選択
- インバースを取り、3Dメッシュの背面頂点を選択



視聴者とV.V.の位置関係



視聴者視点での  
選択メッシュ範囲



視界面選択



視界外選択

# リアルタイム一面3Dデータ生成



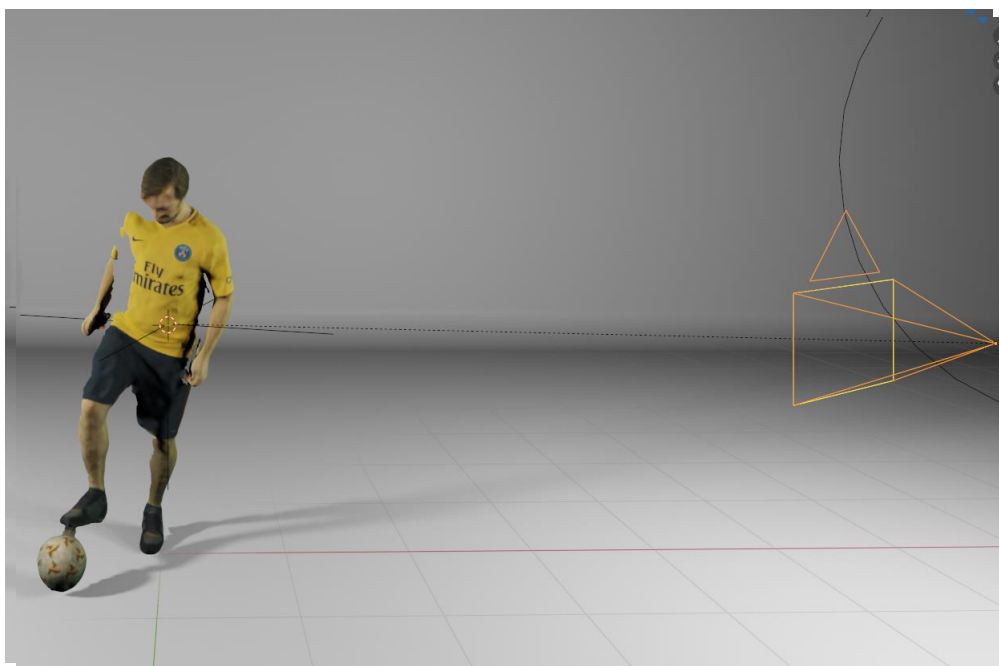
3D空間  
位置取得

V.V.視聴範囲の  
選択

リアルタイム  
一面3Dデータ  
生成

データ圧縮

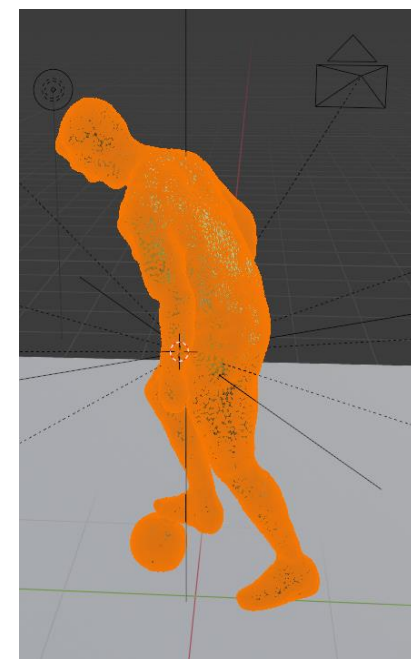
- 選択された視聴者視界背面の頂点を削除
- 視聴者の位置が変更した場合は、リアルタイムに新しい位置に基づいた背面頂点削除



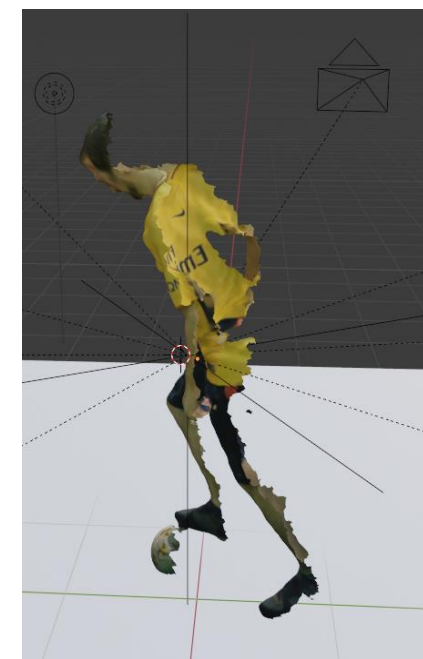
視聴者と一面3Dデータの位置関係



視聴者視点での  
一面3Dデータ



元データ(削減前)



提案手法

一面3Dデータを斜めから見た様子

# Step4 データ圧縮

3D空間  
位置取得

V.V.視聴範囲の  
選択

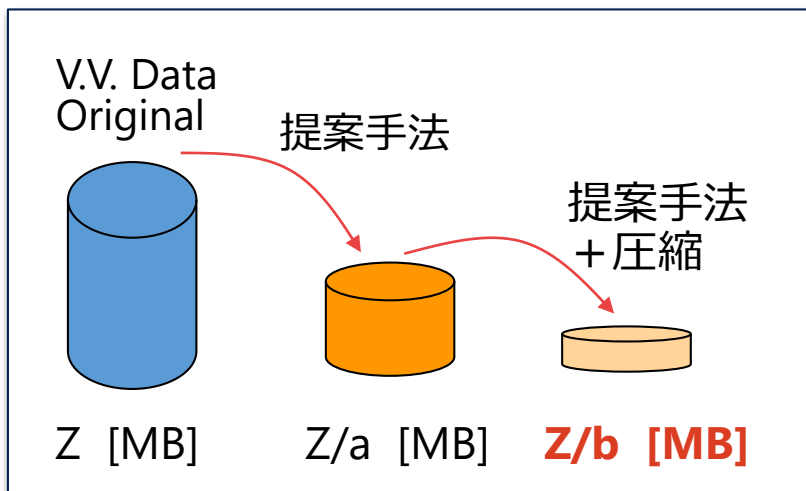
リアルタイム  
一面3Dデータ  
生成

データ圧縮

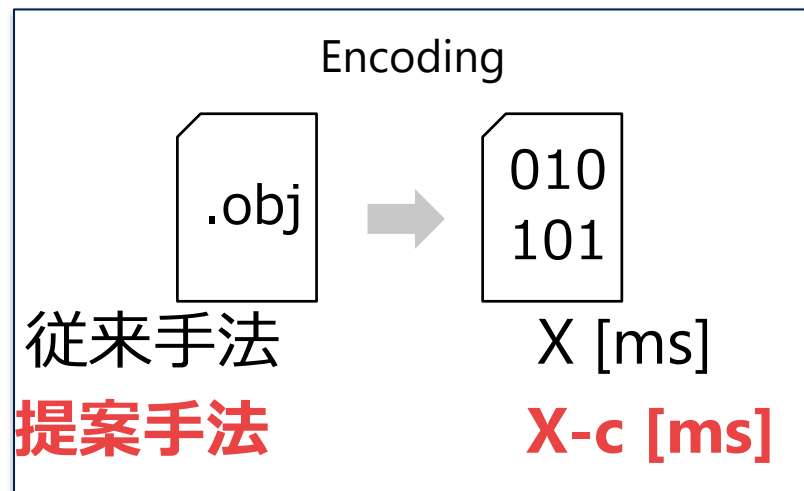
提案手法に圧縮手法[draco]を組み合わせることで一面データをさらに圧縮

- データ量の更なる削減
- サーバー側のエンコード時間短縮
- クライアント側のデコード時間短縮

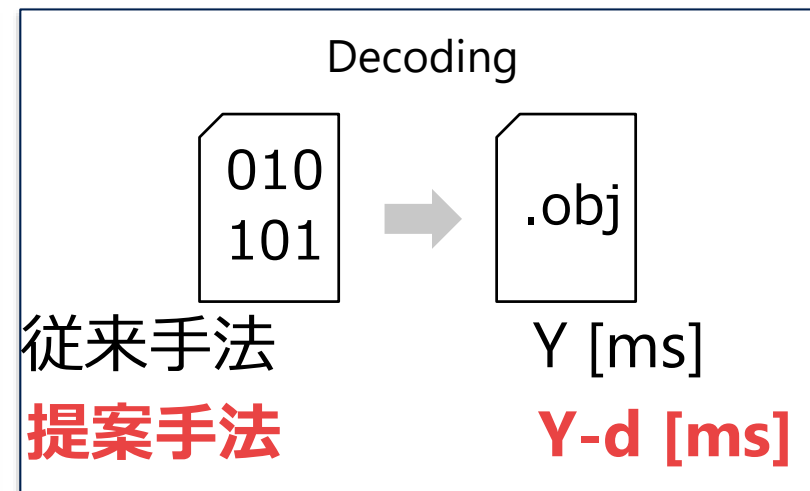
■ 元データ ■ 提案手法 ■ 提案手法圧縮



全体データ量削減圧縮のイメージ



エンコード時間の短縮のイメージ

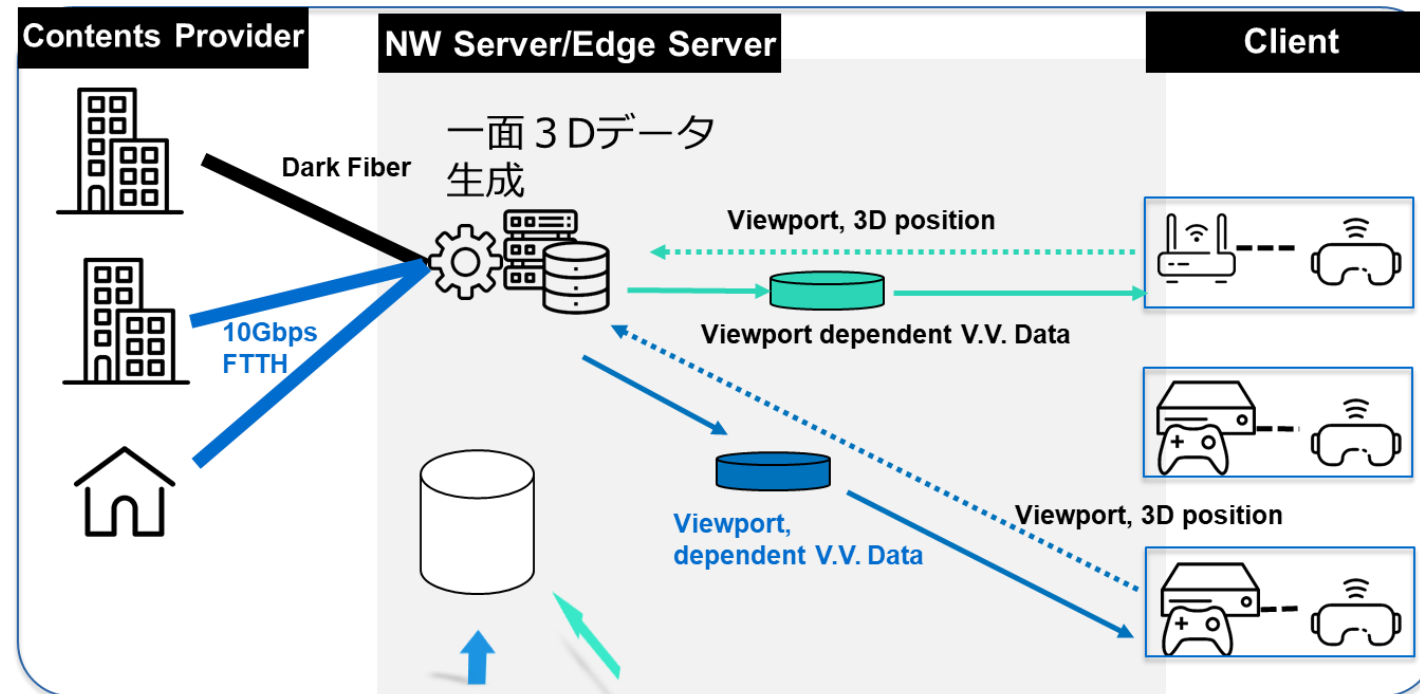


デコード時間の短縮のイメージ

# 配信NWプラットフォーム

提案手法は配信NWプラットフォーム(PF)として組み込まれ、Volumetric Video配信に適用

- **コンテンツプロバイダー:**  
当配信NWプラットフォームへダークファイバーや10G FTTH回線で大容量Volumetric Videoを伝送
- **NWサーバー・エッジサーバー:**  
視聴者情報に基づいて一面3Dデータをリアルタイム生成し、伝送
- **視聴者:**  
FTTH・5G 回線でVolumetric Video(V.V.)を高品質V.V.の複数同時視聴



# 評価実験

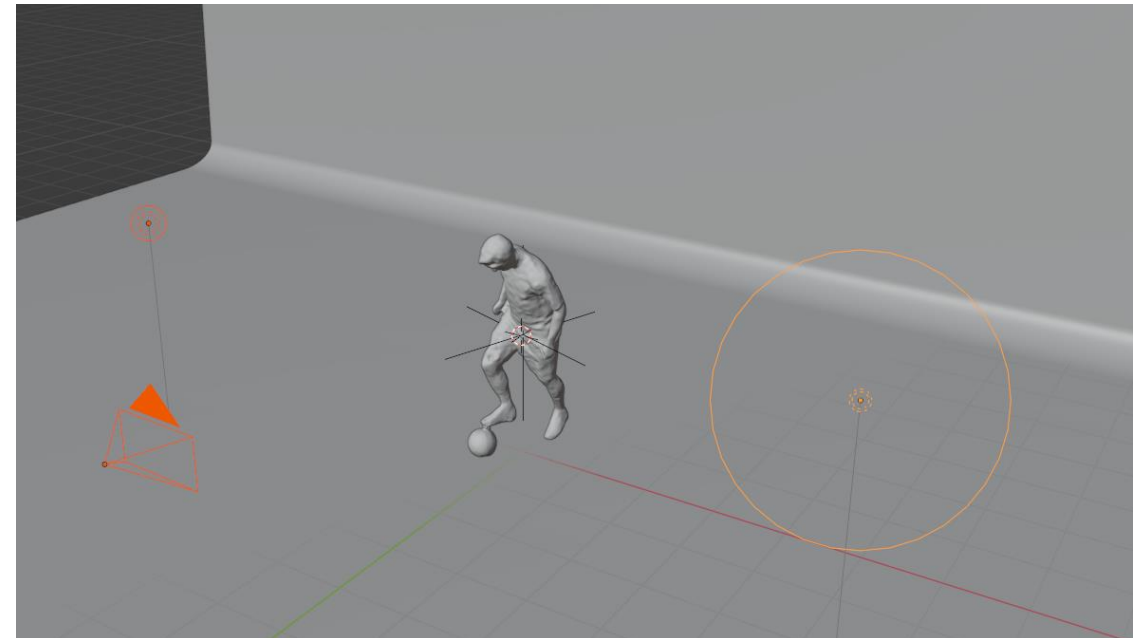


# 実験概要

- **目的：**  
立体空間位置に基づいたリアルタイムデータ削減手法の効果と品質を調べる
- **データセット：**  
[Zerman]のmattisデータの1フレーム
- **実験CG環境：**  
左右前方にポイントライト配置  
カメラ位置は対象Volumetric Videoの頭と同じ高さ



Mattis [Zerman]

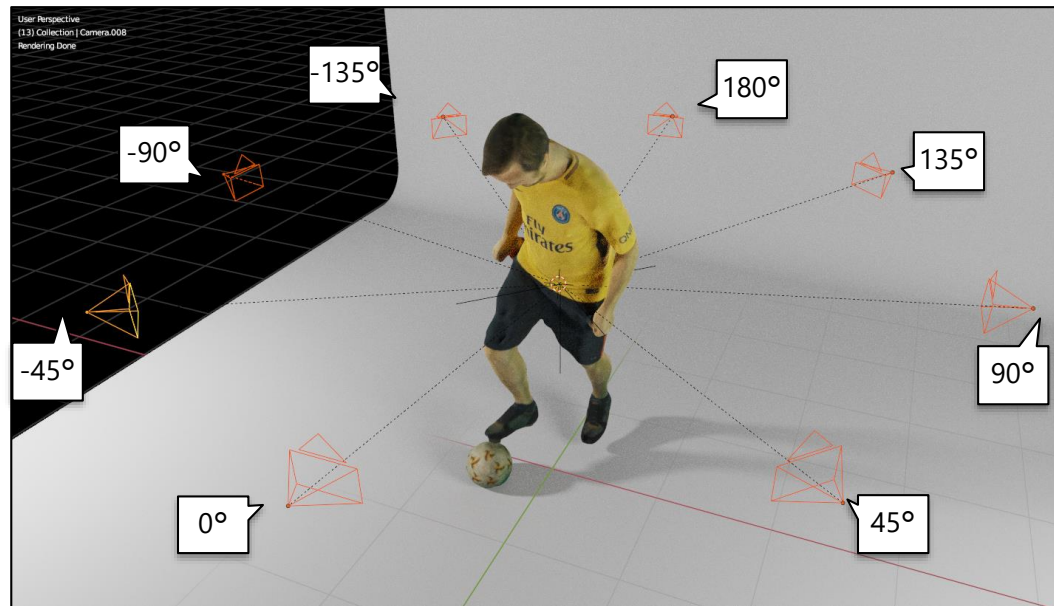


実験CG環境の構成

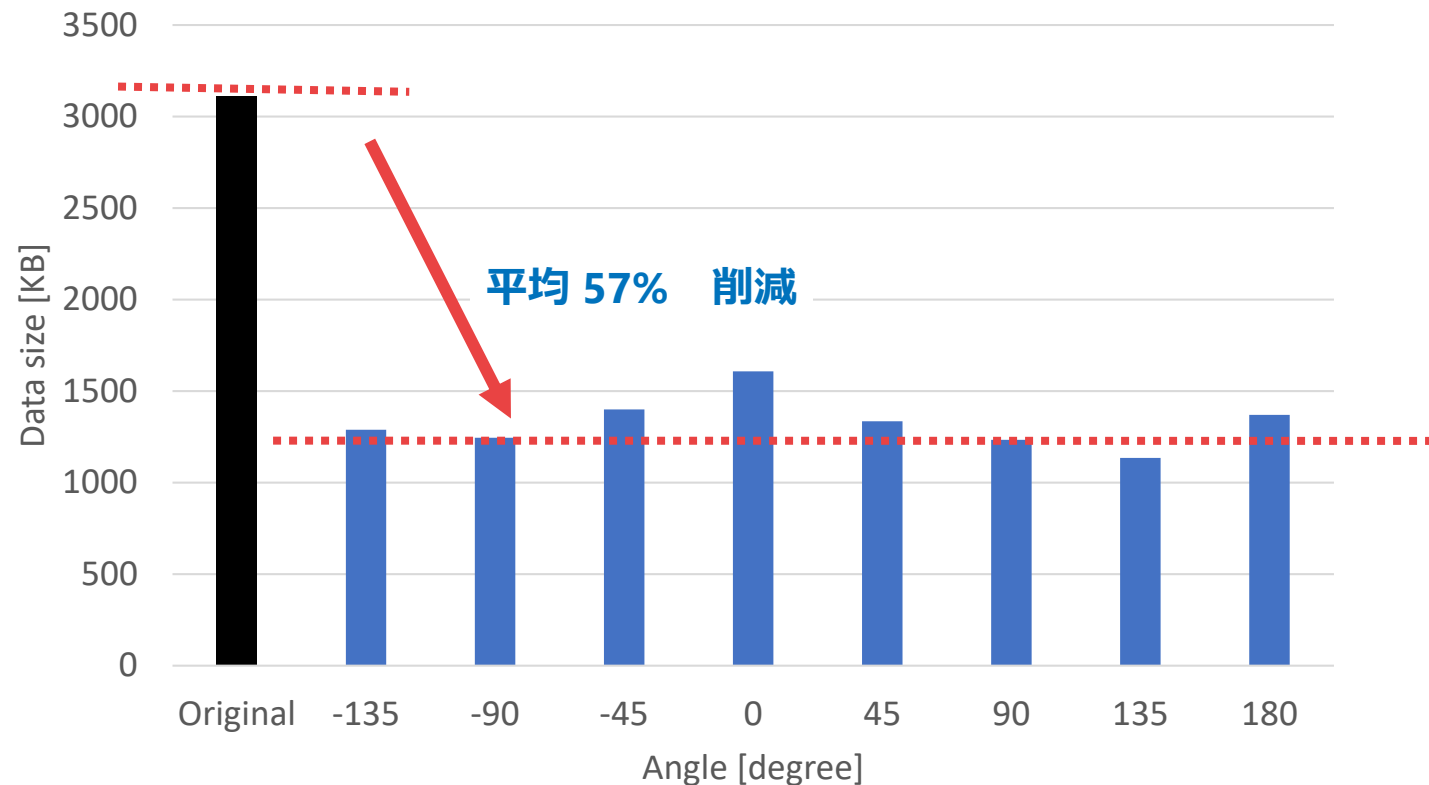
# データ削減量

- 対象データに対し，視聴者視点8つを想定
- それぞれのカメラ位置から提案手法(背面頂点削除)を適用
- 元データと比べ平均57%データ量が削減

データ削減量
品質劣化度
圧縮への効果



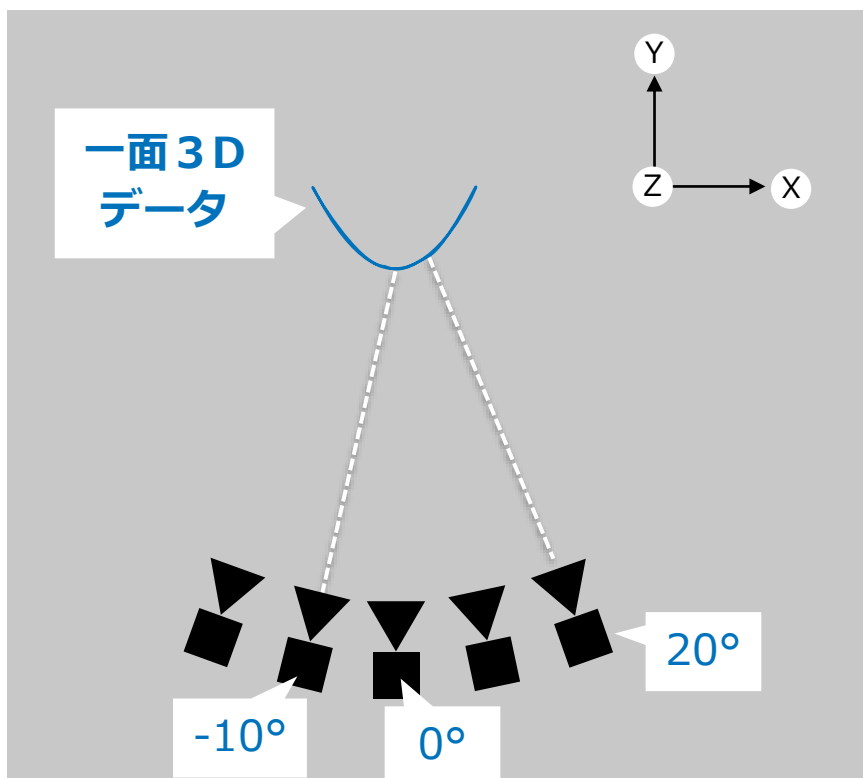
Volumetric Video データ分割での8カメラ方位



提案手法を8方位で適用後のデータ量

# 品質劣化評価

- 視聴者が移動を行い、それに対応する一面3Dデータが遅れて到達した場合を想定
- 元データと一面3Dデータを $-20^{\circ} \sim +20^{\circ}$ の $10^{\circ}$ ずつレンダリング



# 定性評価

~一面V.V.の他角度からの視聴~



-20°



-10°



0°



10°



20°

# 定性評価

- 削除された背面頂点との輪郭部がギザギザ、不連続である
- 近接視聴時には品質劣化があると感じてしまう
- 10°以上ずれた視点での一面3Dデータは、品質劣化が大きい



-20°



-10°

# 定量評価

元データの同一角度映像との差分を2つの品質評価手法で比較

○画素単位比較 Mean Square Error(MSE)

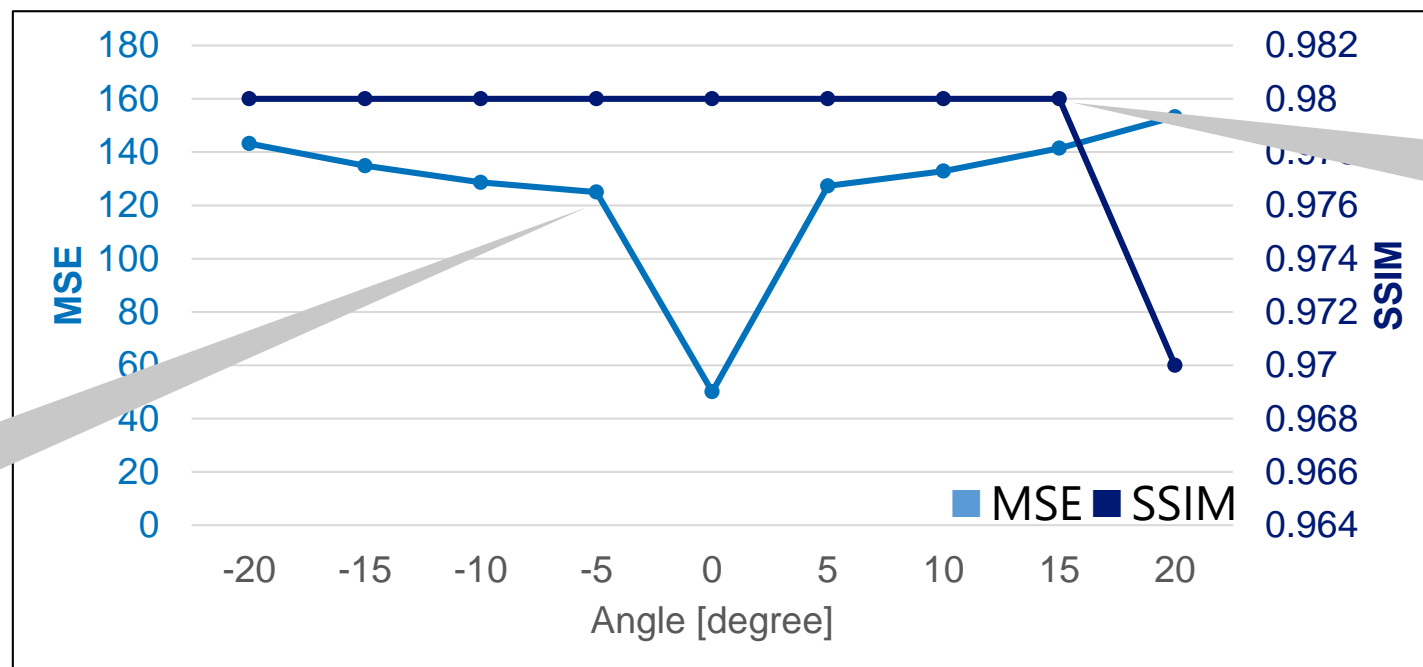
○構造単位比較 Structure Similarity(SSIM)

⇒5° ずれた視点での視聴でも画素単位に品質劣化が見られる。

値：0~  
最高品質時:0

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2.$$
$$\text{SSIM}(x, y) = \frac{(2\mu_x \mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

値：0~1  
最高品質時：1



5°を超えると  
品質の劣化が  
見られる

構造上変化は  
感じにくい。

# 圧縮によるデータ削減効果

提案データ削減手法と圧縮を組み合わせで得られるデータ量縮小効果を調べる

- **実験環境：**

CPU: i7-10750H, メモリー: 32GBのPC, 圧縮手法[draco]を使用

- **圧縮前元データ比較：**

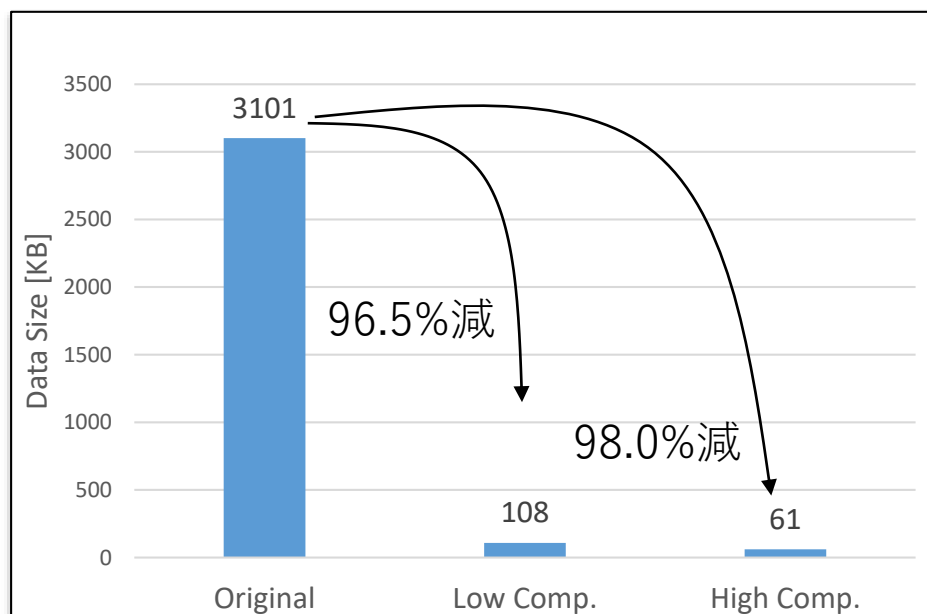
提案手法は元データに対して

低圧縮時に98.0%削減, 高圧縮時に98.9%削減

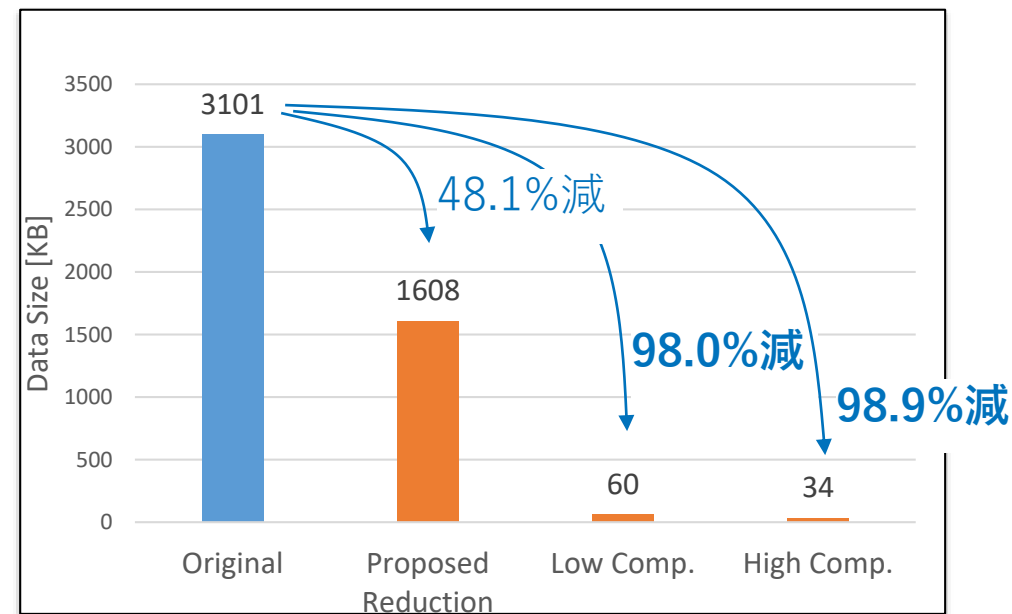
データ削減量

品質劣化度

圧縮への効果



従来手法[draco]

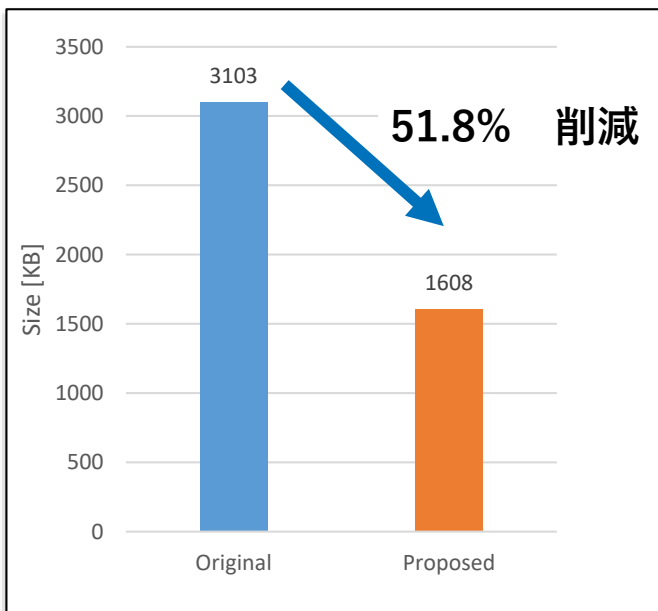


提案手法 (提案削減手法 + 従来手法[draco])

# 圧縮によるデータ削減効果

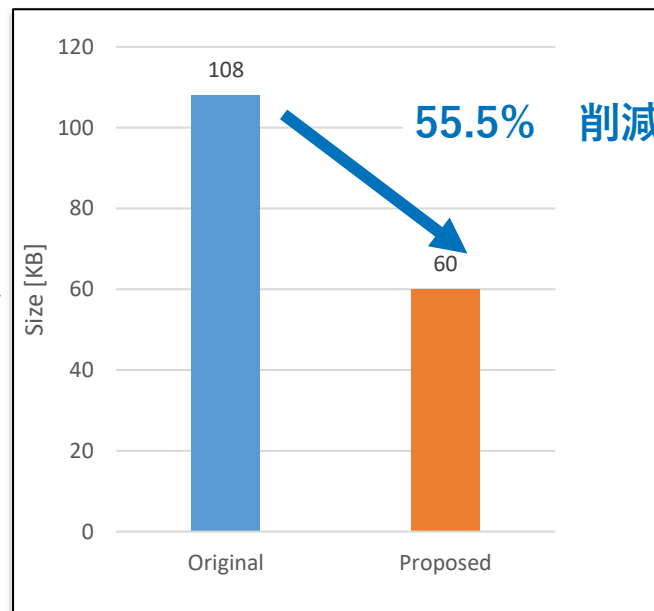
## 圧縮後のデータ量比較：

- 低圧縮時は，元データ圧縮後と比べ**55.5%**のデータ削減(108KB⇒31KB)
- 高圧縮時は，元データ圧縮後と比べ**55.7%**のデータ削減(61KB⇒22KB)

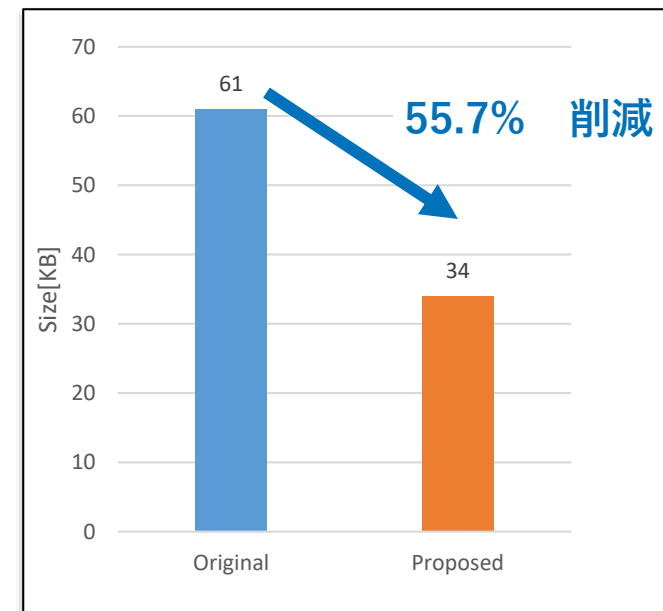


圧縮前データ量  
元データ/提案手法

圧縮



低圧縮後データ量  
元データ/提案手法



高圧縮後データ量  
元データ/提案手法

\*Lower is Better



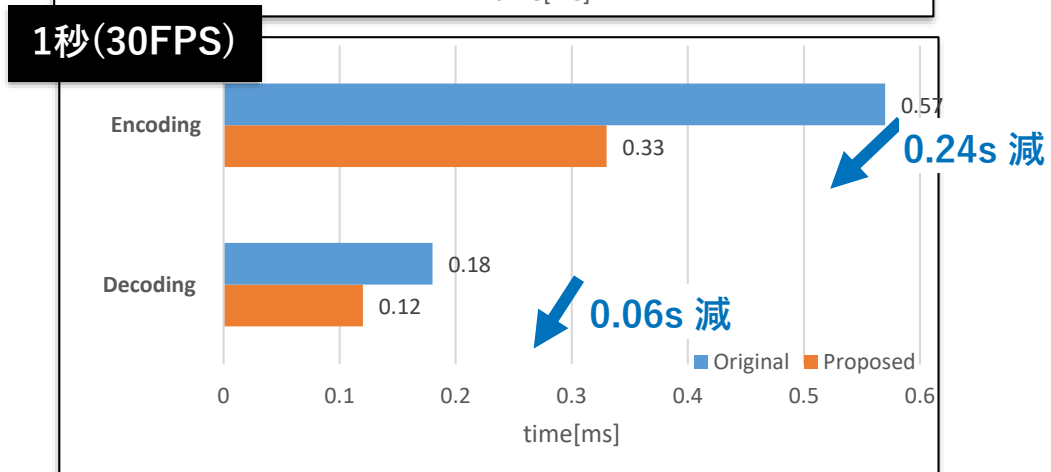
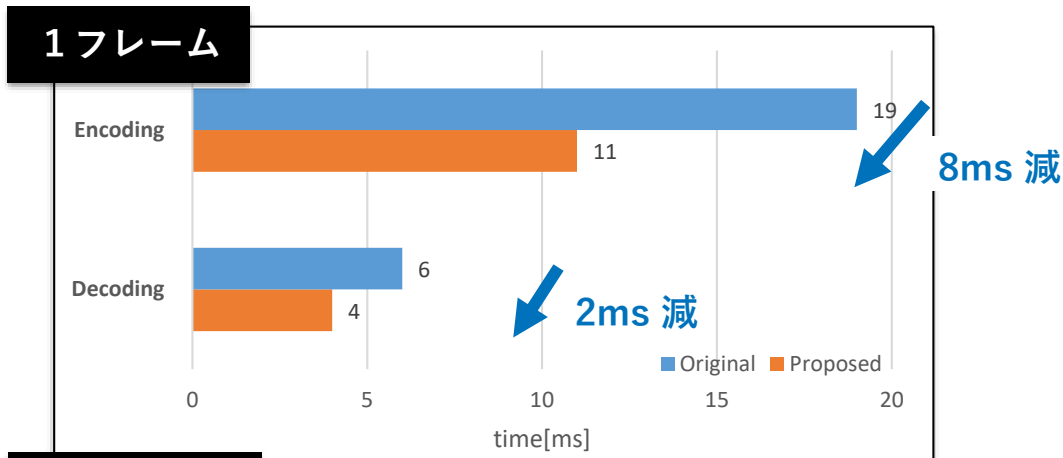
# エンコード・デコード時間の削減

- 提案手法でのデータは、圧縮時のエンコード・デコード時間を短縮
- 低圧縮時サーバーでのエンコード負担を最大8ms(42%)軽減
- 低圧縮時クライアントでのデコード時間を2ms(33%)軽減

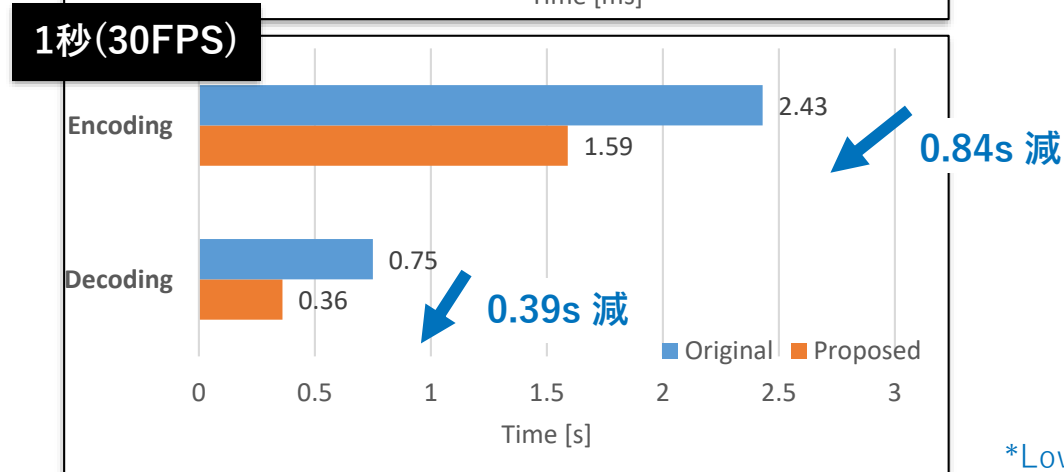
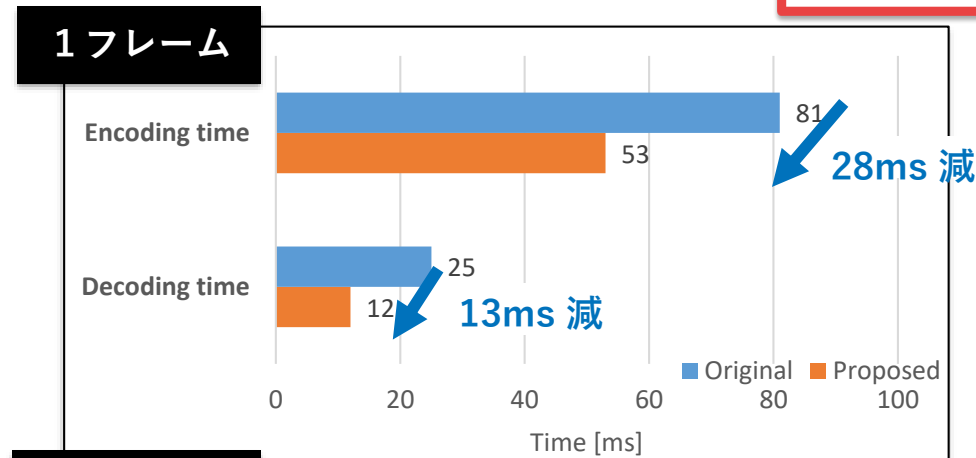
データ削減量

品質劣化度

圧縮への効果



低圧縮時処理時間(元データ/提案手法)



高圧縮時処理時間(元データ/提案手法)

\*Lower is Better

# 考察

提案データ削減手法は、

- 品質を維持したままVolumetric Videoデータを
  - **非圧縮時：平均57%削減**
  - **圧縮時でも55%削減**
- **圧縮エンコード，デコード時間：30%以上短縮**  
これによりトータルの処理時間を削減できる可能性あり

## 今後の検討

- ユーザー数ごとのサーバーでリアルタイム一面データ生成が必要となるため、サーバー処理負担大きく、手法の改善が今後必要
- 背面頂点削除手法は、背面と接続する輪郭部がギザギザしており、手法の改善が今後必要。
- 背面頂点削除にかかる計算コストと時間は、今後調査が必要。

# 考察 (NW観点)

## NW回線で必要となる通信帯域

30FPSのコンテンツ時：

- 非圧縮伝送

800Mbps回線必要 ⇒ **400Mbps回線でVolumetric Videoのメッシュ伝送が可能に**

(テクスチャは+240Mb/s必要、圧縮比率などによりさらにデータ縮小可能) ⇒ 今後検証

- 低圧縮伝送

30Mbps回線必要 ⇒ **20Mbps回線でVolumetric Videoのメッシュ伝送が可能に**

エンコード・デコード遅延があり、Volumetric Video配信と双方向コミュニケーションにおいて全処理遅延150ms以下<sup>[3GPP]</sup>を満たせるかは高性能MECなどを使用し今後検証

Data size	[MB/frame]	[Mb/s](30fps)	[MB/frame]	[Mb/s](30fps)
Before comp.	3.103	744.72	1.608	385.92
Low comp.	0.108	25.92	0.06	14.4
High comp.	0.061	14.64	0.034	8.16

元データ  
データ量・必要伝送帯域

提案データ削減後

Encoding time	[ms] 30fps	[ms] 30fps
Before comp.	NONE	NONE
Low comp.	1.35	0.33
High comp.	2.34	1.59

エンコード時間 (1秒あたり)    デコード時間 (1秒あたり)

Decoding time	[ms] 30fps	[ms] 30fps
Before comp.	NONE	NONE
Low comp.	0.18	0.12
High comp.	0.75	0.36

# まとめ

この発表では、

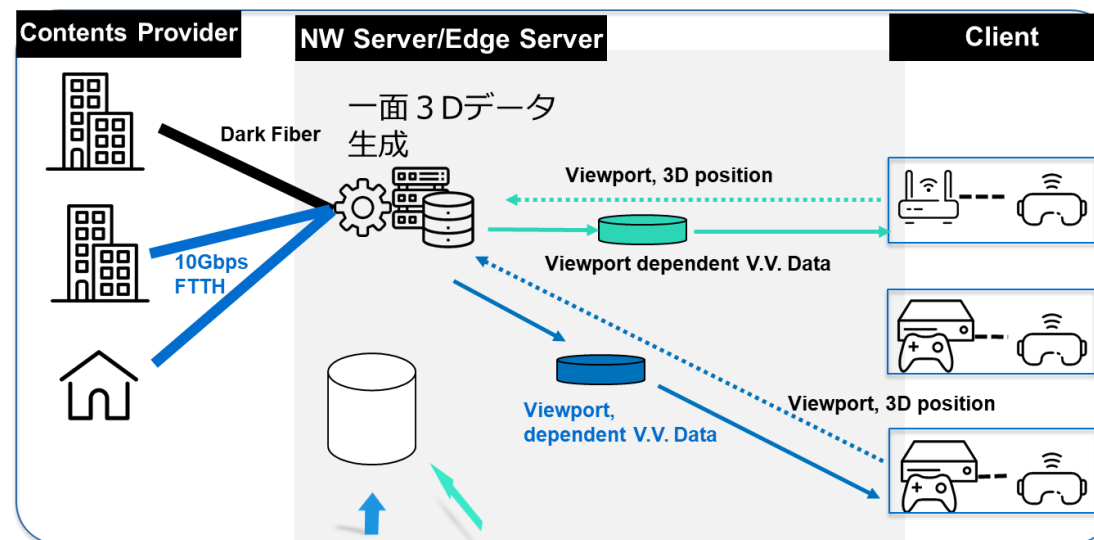
- Volumetric Videoの技術背景
- 視聴者情報に基づいた配信データ削減圧縮手法の提案  
⇒ 従来の57%のデータ量でVolumetric Videoの配信が可能であることを示した

今後は、

- サーバー負荷が少ないデータ削減手法の検討
- 消費者にはアクセス回線でスポーツや番組などのV.V.配信とユーザー間の交流会議サービスへの技術展開



## 将来のVolumetric Videoサービス展開



Source: <http://volumetric-video.com/volumetric-video-calls-using-ef-eve-live-streaming/>, [https://www.youtube.com/watch?v=Ho\\_8UvrAvNs](https://www.youtube.com/watch?v=Ho_8UvrAvNs), <https://www.youtube.com/watch?v=DkOKrVV3SS0>

***Your Value Partner***