自由視点 AR ストリーミングの効率的配信に向けた可視オブジェクトフィルタゲートウェイの実装

河村 侑輝　蛭間　信博　今村 浩一郎

NHK 放送技術研究所 〒157-8510 東京都世田谷区砧 1-10-11

E-mail: {kawamura.y-du, hiruma.n-ho, imamura.k-gu}@nhk.or.jp

あらまし　AR（Augmented Reality）技術を活用して映像メディアの表現空間を 3 次元へ拡張する新たな視聴体験の実現が期待されている．AR 視聴端末へのコンテンツ伝送は無線通信（Wi-Fi，4G/5G）が主となるため，高ビットレートの 3 次元コンテンツの安定受信には多くの課題がある．本稿では，可視オブジェクトフィルタ機能のネットワーク上のゲートウェイへの実装について報告する．可視オブジェクトフィルタ機能とは，6DoF（6 Degrees of Freedom）の空間に配置される出演者や静止物など複数オブジェクトで構成される 3 次元コンテンツのストリーミング伝送において，無線伝送区間の帯域利用を効率化しより安定した受信を可能とする目的で，自由視点 AR の視野内にあるオブジェクトのデータだけを透過させるものである．

キーワード　イマーシブメディア，自由視点 AR，可視オブジェクトフィルタ

An Implementation of Visible Object Filter Gateway for Effective Streaming of Free-Viewpoint AR Content

Yuki KAWAMURA　Nobuhiro Hiruma　and　Koichiro Imamura

Science and Technology Research Laboratories, NHK　1-10-11 Kinuta, Setagaya-ku, Tokyo, 157-8510 Japan

E-mail: {kawamura.y-du, hiruma.n-ho, imamura.k-gu}@nhk.or.jp

Abstract　AR (Augmented Reality) technology is expected to realize a new viewing experience by extending the representation space of visual media to three dimensions. Since content transmission to AR viewer devices is mainly through wireless communication (Wi-Fi, 4G/5G), stable reception of high-bitrate 3D content poses many challenges. In this paper, we propose a visible object filter function implemented in the gateway on the network. In order to improves the efficiency of bandwidth utilization in the wireless transmission and to enable more stable reception in the transmission of 3D content, the visible object filter only passes the data of objects in the field of view of AR among multiple 3D objects such as performers and static objects.

Keyword　Immersive media，Free-viewpoint AR，Visible object filter

1. はじめに

8K 放送など定量的なスペック向上に留まらない映像メディアのさらなる進化と発展を目指し，視聴者に提供する体験の質を高める技術革新が模索されている．その一例として，AR/VR（Augmented Reality and Virtual Reality）技術の活用により，映像メディアの表現空間を 3 次元へと拡張する新たな視聴体験の実現が期待されている．筆者らは，2 次元のテレビ放送を当面の映像メディアの基盤としながら，今後期待される AR/VR 技術の進化や視聴端末の低廉化に従い，徐々に 3 次元メディアが日常生活に広く浸透していくような社会実装プロセスを想定し，テレビ映像との同期提示が可能な3次元コンテンツの伝送方式[1]の研究開発を進めている．本稿では，出演者や静止物など複数のオブジェクトから構成される 3 次元コンテンツの配信技術として，オブジェクト単位でそれぞれ独立した 3 次元データを多重化してストリーミング伝送するオブジェクトベース伝送[2]における，可視オブジェクトフィルタ機能のネットワーク上のゲートウェイへの実装について報告する．可視オブジェクトフィルタ機能は，3 次元空間に出演者や静止物など複数のオブジェクトを配置して 6DoF（6 Degrees of Freedom）の自由視点 AR で視聴可能な 3 次元コンテンツのストリーミング伝送において，AR の視野内にあるオブジェクトのデータだけを透過させる機能であり，無線伝送区間の帯域利用を効率化し，より安定した受信を可能とする．

本稿では，まず，サービス・コンテンツの想定，オブジェクトベース伝送の特徴，可視オブジェクトフィ

ⓒ2021 Information Processing Society of Japan
ⓒ2021 ITE

ルタの実装について述べる．その後，開発した実機による簡易な動作確認結果について述べる．

2. サービス・コンテンツの想定

想定するサービス形態は，図 1 に示すように，テレビ放送による 2 次元の高精細映像と，この映像に同期する 3 次元コンテンツを通信（インターネット）で伝送し，視聴者にそれらをシームレスに組み合わせた新しい視聴体験を提供する．テレビ放送の 2 次元映像と通信経由の 3 次元コンテンツは時間軸・ストーリー展開を共有しており，番組のシーンに応じた 3 次元コンテンツが，放送と同時に配信される．筆者らが提案する 3 次元コンテンツの伝送は，コンテンツ視聴前にデータダウンロードを必要としないストリーミング方式である．視聴者は，番組視聴と同時にスマートフォンやタブレット型端末などのモバイル端末の受信アプリを開いて 3 次元コンテンツを受信し，テレビの前に AR で提示させることで，番組の世界がテレビ画面の枠を越えて 3 次元に拡張された 6DoF の自由視点の視聴体験を楽しむことができる．

テレビ番組の付加サービスとして AR 視聴が可能な 3 次元コンテンツの提供は既に実用化されているものもあるが，放送とは異なる時間軸でオンデマンド提供される関連コンテンツや，放送と同期するコンテンツであっても視聴者に事前にデータダウンロードが要求される．本稿で提案するサービス形態では，データダウンロードの手間など AR コンテンツ視聴における視聴者のインタラクションコストを下げ，2 次元映像と同じように 3 次元コンテンツに日常的に触れられる環境の実現を目指している．

3. オブジェクトベース伝送

提案するオブジェクトベース伝送では，コンテンツを構成する出演者や背景物などのオブジェクトをそれぞれ独立した 3 次元モデルとして符号化して，パケットレベルで識別が可能な形で伝送することが特徴である．パケットのヘッダ部に記載されるパケット識別子（Packet ID）がオブジェクトと対応付けられ，特定のオブジェクトのデータだけを破棄する・抽出するなど対象のオブジェクトを特定した処理を，伝送路上のネットワークノードや受信端末の入力部におけるパケットレベルの処理として容易に実現できる．

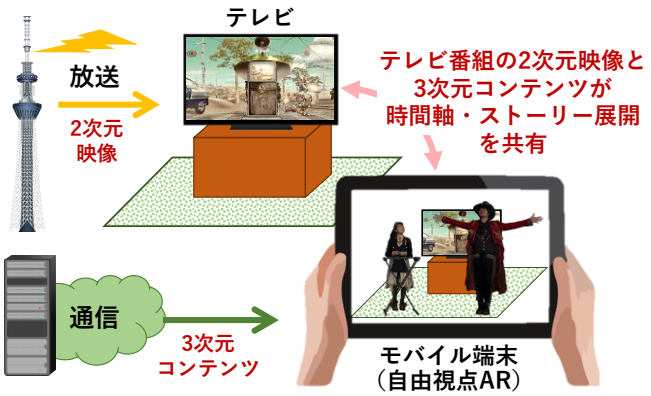

図 1　想定するサービス形態

伝送する 3 次元モデルは，メッシュジオメトリとテクスチャ画像で符号化され，メッシュジオメトリの圧縮として Google Draco[3]，テクスチャ画像圧縮として JPEG，伝送フレーム形式として glTF（GL Transmission Format）2.0 のバイナリ形式である GLB File Format[4] をそれぞれ採用した．各オブジェクトの 3 次元モデルを毎秒 30 フレーム送信し，モバイル端末等の AR 視聴端末でレンダリングし，動く被写体として提示する．

各オブジェクトには，それらを 3 次元空間に配置するための提示位置座標・回転角度などのメタデータを付加する．受信端末は，この空間配置のメタデータに従って各オブジェクトを 3 次元空間上に配置し，実写空間と合成して自由視点 AR で提示する．

4. 可視オブジェクトフィルタの実装

モバイル端末向け OS（iOS，Android）の AR 支援機能（ARkit，ARCore）を活用したアプリケーション実装の利便性から，現状は，スマートフォンやタブレットが AR 視聴端末の中心である．将来的には，AR グラスのような，より軽量・ウェアラブルな端末へと進化

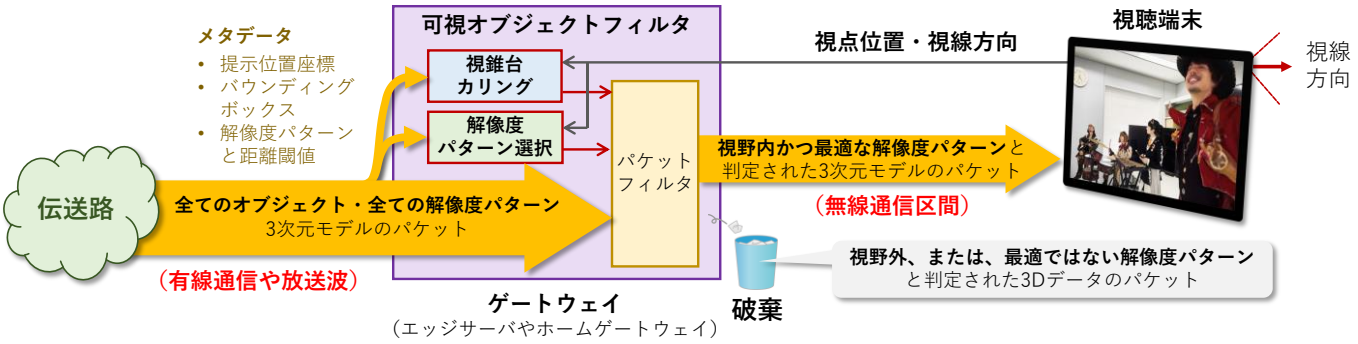

図 2 可視オブジェクトフィルタを含む伝送システム

ⓒ2021 Information Processing Society of Japan
ⓒ2021 ITE

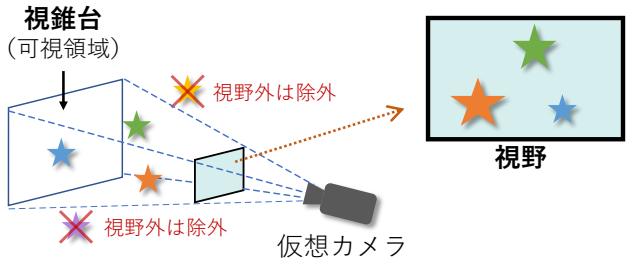

図 3　視錐台カリング

していくことが予想される．こうしたモバイル・ウェアラブル端末へのコンテンツ伝送は無線通信（Wi-Fi や 4G/5G）が主となるため，高ビットレートの 3 次元コンテンツの安定受信が大きな課題となる．

一般に，無線 IP 通信ではトラフィックの輻輳，ノイズや干渉波などにより通信の品質が低下するとパケットロスが増加し，フレーム提示のカクツキや画像破綻などコンテンツの提示品質の低下原因となる．端末 1 台当たりの伝送ビットレートを抑制すれば，接続端末が増えた場合の輻輳を防ぐ効果も期待できる．

提案する可視オブジェクトフィルタは，信号源である配信装置から視聴端末までの通信経路上に設置されるネットワークノード上の処理として実装し，図 2 に示すように，視聴端末からフィードバックされた自由視点 AR の視野状況に応じて，視聴端末に中継するデータを取捨選択する．可視オブジェクトフィルタの上流は有線通信や放送波など十分な伝送容量を確保して IP マルチキャストや CDN（Content Delivery Network）技術を用いて大容量コンテンツの一斉送信が可能なネットワークを，可視オブジェクトフィルタの下流は 4G/5G 通信や宅内 Wi-Fi のネットワークを想定する．可視オブジェクトフィルタを実装するネットワークノードとしては，これらの上流ネットワークと下流ネットワークの境界に位置するエッジサーバやホームゲートウェイなどのゲートウェイを想定する．可視オブジェクトフィルタ処理により視聴端末側に中継するデータを選別することで，4G/5G のモバイル通信や，宅内環境であれば無線 LAN などの無線通信区間のビットレートを低減し伝送を効率化するとともに視聴端末の処理負荷を低減し，より安定したデータ受信とコンテンツ再生を可能とする．

AR の視野状況に応じたデータの取捨選択の手法として，具体的には，視錐台カリングと解像度パターン選択の 2 つの仕組み導入する．

4.1. 視錐台カリングを用いたパケットフィルタ

3 次元コンテンツの視聴において，スマートフォンやタブレットなどの矩形モニタ上にレンダリングされる 3 次元空間の視野は，図 3 に示すようにモニタ中央の手前に位置する仮想カメラの中心点からモニタの 4 頂点を結んで延長した 4 直線に囲まれる四角錐となり，この領域を視錐台と呼ぶ．視野角の広がり具合を一般に FOV（Field of View）と呼び，仮想カメラの焦点距離に相関するパラメタである．視錐台から外れるオブジェクトはモニタ上に提示されないため，視聴端末はデータを受信してもレンダリング処理を省略することができる．視野外のオブジェクトを判定し，処理対象から除くことを視錐台カリングと呼び，CG レンダリングの端末内部処理としては一般的な手法である．

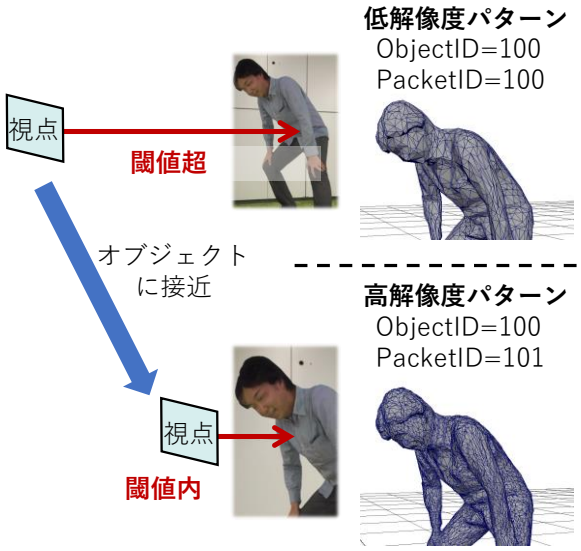

図 4　解像度パターン選択

一方，提案する可視オブジェクトフィルタでは，視錐台カリングを視聴端末内での処理ではなく，視聴端末までの通信経路上のネットワークノードの処理として実装する．視聴端末から可視オブジェクトフィルタに対して，視点位置・視線方向・FOV 角度からなる視野状況をフィードバックして，各オブジェクトについて視野内外の判定を行う．

視錐台カリングの視野内外判定を軽量に実行できるようにするため，各オブジェクトのバウンディングボックス情報をオブジェクトに紐づくメタデータとして送信する．これにより，可視オブジェクトフィルタでは実体の 3 次元モデルを解析することなく，メタデータとして与えられるバウンディングボックスと視錐台の交差判定による視野内外判定が可能となる．

4.2. 3 次元モデルの解像度パターン選択

3 次元空間をレンダリングするモニタ上には，視点位置から遠方のオブジェトは小さく表示され，近傍のオブジェクトは大きく表示される．このため，遠方のオブジェクトは低解像度の 3 次元モデルであってもモニタ上には十分な画質で提示が可能であり，近傍のオブジェクトになるほど高解像度の 3 次元データが求め

ⓒ2021 Information Processing Society of Japan
ⓒ2021 ITE

図 4　実験コンテンツ

られる．予め複数パターンの解像度の 3 次元モデルを用意しておき，レンダリングの提示サイズに応じて使い分けることでレンダリングの処理負荷を低減する手法は，一般に LOD（Level of Detail）処理と呼ばれる．可視オブジェクトフィルタでは，各オブジェクトについて解像度が異なる複数パターンの 3 次元モデルを並列でストリーミング受信し，可視オブジェクトフィルタの視錐台カリングを通過した視野内のオブジェクトについて，さらに最適な解像度パターンを選択して中継する処理として実装する．具体的には，視聴端末からフィードバックされた視点位置と，視野範囲内の各オブジェクトとの空間距離を算出し，設定された距離の閾値に応じて解像度パターンを選択する．

同一のオブジェクトで異なる解像度パターンの 3 次元モデルは共通の Object ID によって紐付けられ，それぞれ異なる Packet ID で多重化されて伝送される．

なお，解像度パターンの構成と，それらを切り替える距離閾値の設定は，視聴品質に影響するパラメタになり得ると考えられるが，本稿では固定的に設定した閾値を用い，切り替え機能の実装と基本的な動作検証に留める．

5. 基本動作検証

検証用コンテンツとして，ボリュメトリックキャプチャスタジオで撮影した 6 名組（ギター，ドラム，グロッケンシュピール，ボーカル，アコーディオン，ウッドベース）のバンド演奏の 3 次元コンテンツを使用する．視聴端末では，図 4 に示すように各演奏者を実空間に重ね合わせて 3 次元空間に配置し，6DoF の自由視点 AR で提示する．各出演者の 3 次元モデルについて，静的領域と動的領域を分離して別オブジェクトとして符号化し，静止領域の送信フレームレートを下げることで可視オブジェクトフィルタの上流でも伝送ビットレートを低減する工夫を行う．ギターとアコーディオンについては演奏者が座る椅子を，ドラムとグロッケンシュピールについては演奏者と楽器を分離して，これらを 4 個の静止オブジェクトとした．

表 1　オブジェクト構成とデータ諸元

オブジェクト	Object ID	Packet ID	頂点数	テクスチャ画素数（正方の一辺）	平均ビットレート（Mbps）
ギター奏者	100	100	2000	512	19.2
		101	8000	1024	41.3
ドラム奏者	200	200	2000	512	24.5
		201	8000	1024	36.0
グロッケン奏者	300	300	2000	512	18.6
		301	8000	1024	38.2
ボーカル	400	400	2000	512	16.0
		401	8000	1024	31.6
アコーディオン奏者	500	500	2000	512	26.6
		501	8000	1024	56.7
ベース奏者	600	600	2000	512	24.7
		601	8000	1024	53.4
椅子①	700	700	2000	512	1.1 (1fps)
ドラムセット	800	800	8000	1024	1.8 (1fps)
グロッケン	900	900	2000	512	1.1 (1fps)
椅子②	1000	1000	2000	512	1.1 (1fps)

合計 391.9 Mbps
（高解像度パターン＋静止オブジェクト 262.3 Mbps）

表 1 にオブジェクト分離の結果，実際に多重され伝送されるオブジェクトの構成と諸元を示す．演奏者 6 人の動的領域の 3 次元モデルは，それぞれ高解像度と低解像度の 2 パターンの 3 次元モデルを毎秒 30 フレームで並列に送信する．4 個の静止オブジェクトについては，毎秒 1 フレームで送信する．これらの合計 16 オブジェクトの 3 次元モデルの送信で，可視オブジェクトフィルタの上流側の時間平均のビットレートは 391.9 Mbps となった．なお，低解像度パターンを除いたビットレートは 262.3 Mbps となり，低解像度パターンを追加的に伝送することによる可視オブジェクトフィルタの上流のビットレートの増加は 129.6 Mbps となる．

また，2 種の解像度パターンを切り替える閾値となる視点位置・オブジェクト位置間の距離を，コンテンツを実物大とした時のスケールにおける 2m と設定して，これをメタデータとして送信する．コンテンツを 1/2 スケールで実空間に AR 提示した場合，実空間上での閾値は 1m に相当する．

可視オブジェクトフィルタは，合計 16 オブジェクトの 3 次元モデルが多重化された平均 391.9 Mbps の信号を受信し，視聴端末からフィードバックされた視点位置・視線方向に応じて，必要なオブジェクトの 3 次元モデルのみを視聴端末に送信する．図 5 に，可視オブジェクトフィルタの動作状況を可視化した画面と，同じタイミングの視聴端末の画面キャプチャを示す．

ⓒ2021 Information Processing Society of Japan
ⓒ2021 ITE

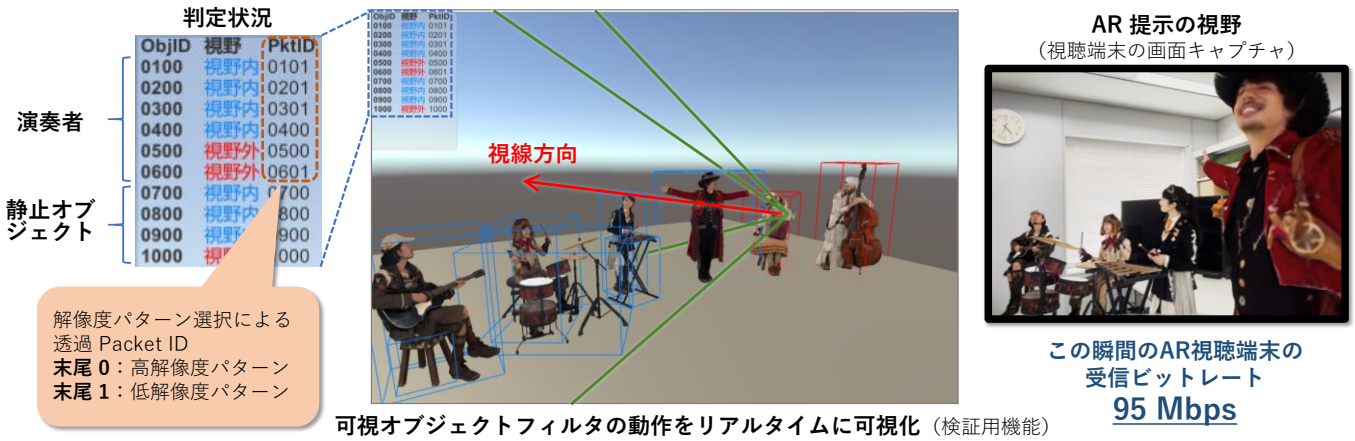

図 5　可視オブジェクトフィルタの動作例

可視化画面はフィルタ機能に必要なものではないが，動作検証用の画面として実装し，コンテンツの 3 次元空間中における視聴端末の位置，視錐台をリアルタイムに表示するとともに，各オブジェクトのバウンディングボックスの色で視錐台カリングの判定状況を確認できる．また，解像度パターン選択の結果については，画面上に表示される通過 Packet ID の一覧（図 5 左上）により確認できる．可視化画面と視聴端末の画面キャプチャを比較すると，アコーディオンとベースは正しく視野外と判定され，解像度パターン選択についても，視点から最近接しているボーカルだけが高解像度パターン，その他の遠方のオブジェクトは低解像度パターンが選択されており，正常に動作していることが分かる．

同じタイミングで測定した瞬間的な視聴端末の受信ビットレートは 95 Mbps であった．時間平均と瞬時値での比較となるが，これは可視オブジェクトフィルタの上流のビットレートである 391.9 Mbps よりも 297 Mbps 低く，低解像度パターンを送信しない場合の時間平均のビットレート 262.3 Mbps と比較しても 167 Mbps 低いビットレートである．複数の解像度パターンの並列送信は，可視オブジェクトフィルタの上流のビットレートを増加させるというデメリットはあるものの，視野状況に応じたフィルタ処理を追加することで，

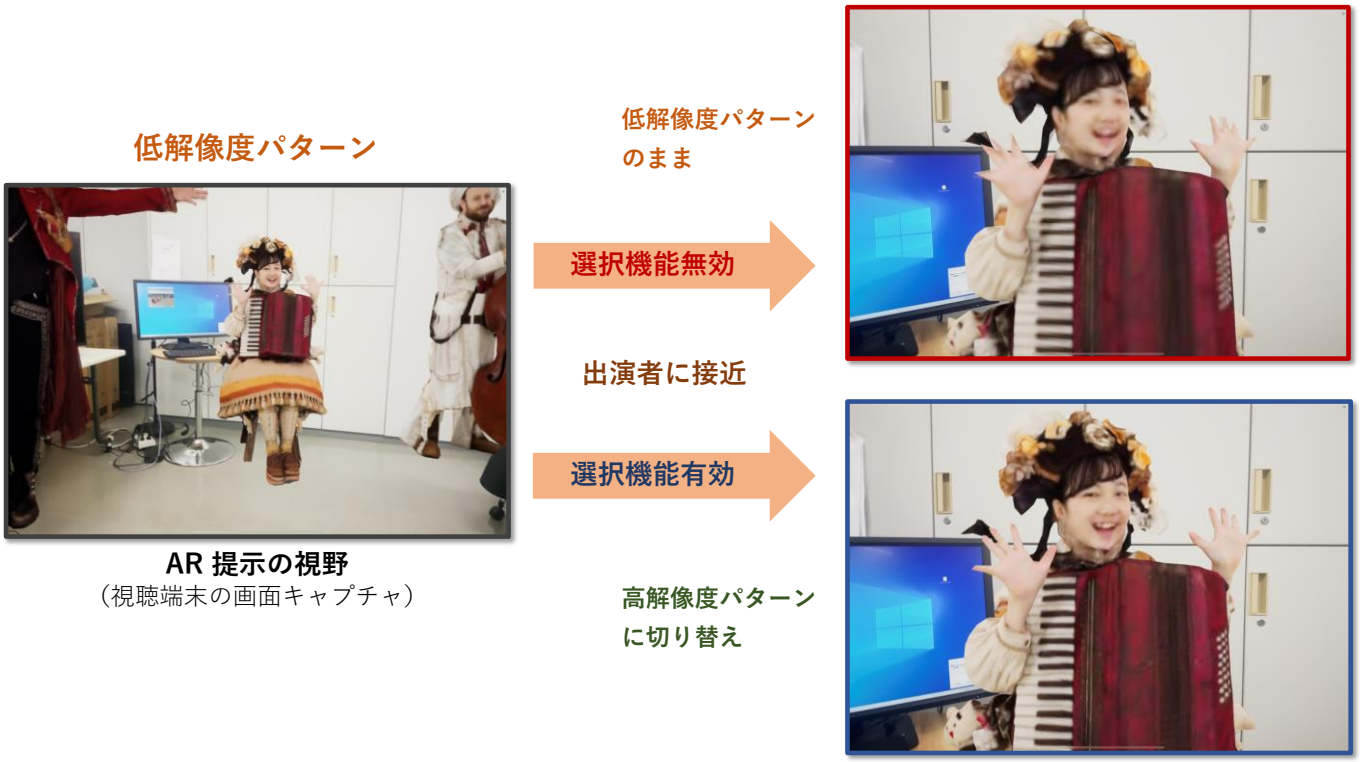

図 6　解像度パターンの切り替えと提示品質の例

ⓒ2021 Information Processing Society of Japan
ⓒ2021 ITE

可視オブジェクトフィルタから視聴端末までの無線区間の伝送の効率化には大きな効果があることが確認できた．

図 6 に，解像度パターンの切り替えによる提示品質の例を示す．演奏者のオブジェクトを等身大で AR 提示し，解像度パターン選択の閾値として設定した 2m よりも離れた視点位置での視聴端末上の提示品質と，2m 以内に接近にした視点位置における視聴端末上の提示品質を視聴端末の画面キャプチャにより比較している．さらに，オブジェクトの 2m 以内に接近後の視聴端末の画面キャプチャの 2 つの画像（図 9 右側の上下）は，解像度パターン選択を無効化した状態で接近して低解像度パターンのまま提示した場合と，解像度パターン選択を有効化して高解像度パターンに切り替えて提示した場合の比較である．低解像度パターンのまま接近すると，指先や髪飾りなどの形状が崩れ，テクスチャも精細感が失われるのに対して，解像度パターン選択を有効化して高解像度パターンに切り替えられた場合には，指先や髪飾りの形状などが維持されテクスチャの精細感が向上していることが分かる．

6. まとめ

本稿では，複数のオブジェクトから構成される 3 次元コンテンツの効率的伝送に向けた，可視オブジェクトフィルタ機能のネットワーク上のゲートウェイへの実装，機能検証について報告した．今後，より実用に近い環境での効果・性能評価を実施する予定である．また，解像度パターンの構成・閾値の設定による視聴品質への影響と設定の最適化に関しても，調査検討を行いたい．また，クラウドレンダリングとの性能比較や，提案した可視オブジェクトフィルタとクラウドレンダリングの双方のメリットを活かしたハイブリッドな方式の検討も行いたい．

文　献

[1] 河村 侑輝，“テレビ映像と同期した自由視点 AR コンテンツのリアルタイム伝送技術の開発，” 映像情報メディア学会誌，vol.75，no.1，p.131-138，Jan. 2021

[2] 河村 侑輝，前田 恭孝，久富 健介，今村 浩一郎，“3 次元コンテンツのオブジェクトベース伝送と複数提示方式による適応的視聴，” 電子情報通信学会技術報告，vol.121，no.73，SIS2021-6，p.32-36，June 2021

[3] Frank Galligan，“Draco Bitstream Specification，” https://google.github.io/draco/spec/，Oct. 2017

[4] Khronos Group，“The glTF 2.0 Specification，” https://github.com/KhronosGroup/glTF/tree/master/specification/2.0/

ⓒ2021 Information Processing Society of Japan
ⓒ2021 ITE