骨格部位毎の変動性を考慮した人間/動物の点群の軽量化とストリーミングを行う3次元空間共有システム

石丸 大稀[1]　中村 優吾[2]　藤本 まなと[3,4]　諏訪 博彦[1,4]　安本 慶一[1,4]

概要：近年，メタバースやデジタルツインへの注目が高まっており，現実世界の空間を仮想空間に取り込み観光分野などに応用する試みがなされている．しかし，現実空間は膨大かつ複雑なため，仮想空間へのプロジェクションは困難な課題である．そのため，シンプルかつ汎用性の高い点群映像への期待が高まっており．現実空間を仮想空間にストリーミングすることで空間共有を実現する．しかし，点群データはサイズが大きく，ネットワークや計算資源に制限のある環境では，リアルタイムでのストリーミングが困難という問題がある．本研究では，ユーザの体感品質（QoE: Quality of Experience）を維持しつつ，低い帯域幅消費で点群のリアルタイムストリーミング・3D 再構成を行う手法を提案する．提案システムでは，現実空間から人や動物のみの点群を取得し，身体部位ごとに品質制御を行うことで送信データ量を削減したのち，4G/5G 通信を介して仮想空間にストリーミングするシステムを提案し，実装・評価を行った．評価実験では，提案システムが出力する点群の品質がユーザの QoE に与える影響を調査するため，提案手法および 5 種類の比較手法（全体的に解像度を減じる手法など）を用いて，ストリーミングされた点群を 3D 再構成した映像を作成し，滑らかさ・解像度・総合品質の 3 項目の MOS （Mean Opinion Score）で評価した．その結果，総合評価では比較対象の５手法のうち３手法に対して提案手法が有意（$p < 0.01$）に優れていることが確認された．また，提案手法と同程度であると評価された残る 2 手法に対しても，滑らかさまたは解像度において有意に優れていることが確認された．

1. はじめに

近年，メタバースやデジタルツインへの注目が高まっており，現実世界の空間を仮想空間に取り込み観光分野等に応用する試みがなされている．しかし，現実空間は膨大かつ複雑なため，仮想空間へのプロジェクションは困難な課題である．現実空間のデジタル化に適した表現形式として，シンプルかつ汎用性の高い点群映像が期待されている．現実空間を点群映像として仮想空間にストリーミングし，3D での再構成を行うことで，遠隔ユーザに対して臨場感のある空間共有が実現できる．

しかし，点群データはサイズが大きく，ネットワークや計算資源に制限のある環境では，リアルタイムでのストリーミングが困難という問題がある．この問題に対し，点群データのエンコーディング/デコーディングやタイリング，視野角予測を用いた伝送最適化など，多くの研究・アプローチが存在する．しかし，これら研究の多くは点群の圧縮やクライアントベースで不要なデータの送信を防ぐことに焦点が当てられている．また，それらの制御がユーザの QoE にどのような影響を与えるのか評価した研究も存在するがその数は多くない．

我々は，人間や動物の点群シーケンスに対して，点群全体の品質（解像度やフレームレート）を一括で制御するシステムをすでに提案している [1]．本稿では，このシステムを拡張し，ユーザの体感品質（QoE: Quality of Experience）を維持しつつ，低い帯域幅消費で点群のリアルタイムストリーミング・3D 再構成を行う手法を提案する．点群をた上でそれぞれの部位に品質制御を適用する．提案システムでは，人間や動物の身体・骨格部位には激しく動く部分とそうでない部分があるという変動性に着目し，体感品質をできるだけ下げないような解像度とフレームレートを骨格部位ごとに設定しストリーミングすることで，使用帯域幅を利用可能帯域幅以下に減らす QoS 制御を行う．現実空間（観光地）の人や動物の点群を 4G/5G 通信（上り帯域 17〜39Mbps[*1]）を介して仮想空間にストリーミングすることで，観光地の空間を遠隔ユーザへ３次元で空間共有するシナリオを想定し，システムの検討・実装を行った．

[1] 奈良先端科学技術大学院大学
Nara Institute of Science and Technology
[2] 九州大学, Kyushu University
[3] 大阪公立大学, Osaka Metropolitan University
[4] 理化学研究所, RIKEN

[*1] 実効速度計測結果（NTT ドコモ）[2]

評価実験では，システムが出力する点群の品質がユーザの QoE に与える影響を調査するため，15 人の被験者に対して，提案手法および 5 種類の比較手法（身体部位の分割なしに，全体の解像度を 10%から 100%（オリジナル）まで，フレームレートを 15FPS から 2FPS まで 5 段階で変更したもの）を用いてストリーミングされた点群を 3D 再構成した映像の主観評価を，滑らかさ・解像度・総合品質の 3 項目の MOS （Mean Opinion Score）を算出することで行った．

その結果，提案手法の MOS は，総合評価で 2.8，なめらかさ 3.7，解像度 2.7 となり，総合評価では比較対象の 5 手法のうち 3 手法に対して有意（$p < 0.01$）に優れていることが確認され，提案手法と同程度であると評された残る 2 手法に対しても，滑らかさまたは解像度において有意に優れていることが確認された．また提案手法を用いて，点群ストリーミングに必要な帯域幅を Gbps レベルからおよそ 20Mbps にまで削減できたことから，4G/5G 通信下でリアルタイム点群ストリーミングが実現可能であることを示した．

2. 関連研究

本章では，バーチャル観光ツアーと XR コラボレーション，点群映像コラボレーションに関する既存研究について述べる．

2.1 バーチャルツアー

バーチャルツアーに関する研究では，360 度映像を活用したものがある．Nassani らは既存のオンライン会議システムを使用して 360 度映像をライブ配信し，それと同時に視聴者の視点（360 度映像の中でどの部分を見ているか）やポインティング（視聴者が映像の中から特定の一点を指定する）を共有することで互いにインタラクションを行うというシステムを提案している [3]．このシステムでは 360 度カメラから得られた映像シーケンスや視聴者の視点情報（視聴者が 360 度映像のどの部分を見ているか・どの部分に興味を示しているか）などを WebRTC を用いて配信者-ユーザ間でリアルタイムに交信し，相互のインタラクションを実現している．360 度映像による自由視点での体験に加え，配信者（司会者）と参加者のインタラクションを通して，本物のツアーに参加しているかのような感覚が得られるだろうと述べており，観光に限らず，教育・研修，商談，視察，公演の鑑賞など様々な領域で活用できるとしている．一方，このシステムではユーザが自らの意思で移動することはできず，自由度としては，3DoF[*2]に留まっている．

[*2] 3 Degree of Freedom，頭・首の回転や傾きに対応するが移動には対応しない．

2.2 XR コラボレーション

現実に近いコラボレーションを実現するため，3 次元情報を用いる例がある．Lee らは，現実空間を関心対象（OOI: Object Of Interest）と環境対象（AO: Ambient Object）に分け，OOI の事前再構成と AO のリアルタイム再構成を組み合わせたアプローチを提案している [4]．OOI については，オブジェクトの特徴情報から検出される空間内での姿勢（位置・方向）情報をリモート側と共有することでポリゴンモデルでのリアルタイム再構成を行っている．AO については，OOI 以外の部分をマスクした点群を生成してリモート側に送信し，リモートコラボレーション時に背景として使用されている．また，OOI の姿勢情報の送信，AO の点群情報の送信には，それぞれ WebSocket と WebRTC が用いられており，サーバへの負荷やデータサイズに応じて各情報の転送手法が選択された．さらに，この論文内では，OOI としてボールバルブを使用する例が示されており，リモート側の有識者から操作方法の指示を得るといったケースにおいて有用性があることが示されている．現実の物理空間と仮想空間を連携し，高い再現性で協働を可能とするシステムを提案しているが，その一方で OOI の認識や事前モデル化といった工程により，利用ケースが限られている．

2.3 点群映像ストリーミング

点群の圧縮は，デコードが複雑であるという問題が存在する．Li らはこの複雑さを解決するため，圧縮されたタイルに加えて異なる品質レベルの非圧縮タイルを送信するアプローチを提案している [5]．このアプローチでは，点群を均等に分割した各タイルに対して，適切な品質レベルを選択し，計算資源と通信資源のトレードオフを考慮した動的制御を行っている．非圧縮のタイルは帯域幅を多く消費するが，デコードの必要がないため計算量を減らすことができる．またタイル選択に関しても，ユーザの FoV（Field of View）内に存在するタイルのみを選択することで不要なデータ送信を削減し，品質の選択と併せて QoE を最適化することができるとしている [6]．そして著者らは，このシステムでの QoE を測定し，従来の方式と比較してより高い QoE 値を得ることができたと示している．点群映像ストリーミングにおける通信・計算資源の配分方法について提案を行い，既存の方式と比べて優れた性能を持つことを示した．しかし，このシステムに用いられた均等タイリングと品質制御には，対象物の動きや文脈が考慮されていない．我々のシステムは上記のアプローチを拡張し，対象物の動きや文脈に応じて点群の品質（つまり解像度やフレームレート）を動的に制御するものである．

2.4 本研究の位置づけ

このように，従来研究では，360 度映像のためユーザが

自由に移動できない（3DoF），OOI の事前のモデル化が必要，品質制御に OOI の動作が考慮されていない，といった課題がある．本稿では，これら３つの課題を解決する方法を提案する．また，リモート観光というシナリオでの仮想化（オンライン化）を実現するために，スマートフォンやセルラーネットワークの利用を前提にする．この条件下での点群ストリーミングは，データ量が大きく，転送を行うためにデータの削減や圧縮を行う必要がある．このため本稿では，これらの条件下で実現可能なシステムを設計・提案する．また，システムの評価としてユーザの QoE を測定し，提案手法がユーザの体感品質に与える影響を評価する．

3. システム要件

本稿では，現実空間のオブジェクトをリアルタイムに３次元スキャン・転送し，仮想空間上で再構成することで，遠隔ユーザと現地ユーザの空間共有を３次元で実現するシステムを提案する．このセクションではシステムの前提条件とそれに伴う問題点を整理し，また我々が想定するユースケースも紹介する．

3.1 前提条件

本システムが想定しているシチュエーションと，その前提となる条件について述べる．まず，我々はシステム利用のシチュエーションとして，観光分野を想定している．これは，観光という行動は現実空間での体験や感覚が重要視されるため，仮想化が難しいユースケースの一つであり，現実空間の人々や動物との触れ合いを遠隔で実現することはあらゆる場面に応用可能であると考えているからである．このタスクを扱う研究は多くなく，チャレンジングな課題である．

そしてこのシチュエーションを想定すると，上記の目標達成のためにいくつか条件が必要になる．条件は大きく２つ存在し，高速なネットワーク回線が利用できない可能性があるという点と固定されたシステムが適切でないという点である．

まず，ネットワーク回線に関する条件について説明する．観光地は，地方から都市部まで幅広く点在し，屋外・屋内を問わない．我々は提案するシステムが多くの場所で幅広く利用可能である必要があると考えている．そのため，セルラーネットワークの利用を前提とする．日本の通信事業者である docomo の資料によると，2022 年 1 月から 3 月時点での日本国内におけるセルラー通信の実効通信速度はダウンロードが 137Mbps〜273Mbps，アップロードが 17Mbps〜39Mbps であることが示されており，光回線には及ばない，そのため，これらの通信速度を考慮してシステムを検討した．

次に，固定されたシステムが適切でないという条件について説明する．我々は現場にセンサや処理装置を固定/設置することは汎用性の点から適切ではないと考えた．その理由は，現実空間のオブジェクトをスキャンする際の対象となるオブジェクトとして人間や動物を想定しているが，彼らがずっとその場に留まるというシーンは少ないと考えられるからである．システムの利用範囲を広げるために，現地ではモバイルデバイスで完結するシステムである必要があるとの結論を得た．この条件から，我々は現地にモバイルデバイスを扱うためのボランティアが必要であると考えた．各観光スポットにボランティアがいる前提で，彼らの協力を得ることでシステムが実現されるとしている．

以上がシステムの実現に際して考慮すべき条件である，我々はこの条件を前提とし，システムの検討を進めた．

3.2 想定するユースケース

本システムは，現地での観光体験を仮想空間上に再現することにより，旅行・観光を更に身近なものとしつつ，旅行者らと現地の人々の双方が利益を享受するシステムの構築を目指している．システムの最終的な利用例として，以下の流れを想定している．

(1) 対象観光地に LiDAR を備えたモバイル端末を持ったボランティアがいる．
(2) 遠隔ユーザは，観光地に対応する仮想空間（予め提供されている）内を自由に移動し，視野に含まれる現実空間の動的オブジェクト（人や動物など）の仮想空間への投影を依頼する．
(3) ボランティアは，依頼を受け，オブジェクトをモバイルデバイスが備える LiDAR でスキャンしシステムに送信する．
(4) オブジェクトが仮想空間内に投影される．

遠隔ユーザと現地オブジェクトの間にオブジェクトのスキャンを担うボランティアを配置することで，ボランティアを介した現地とのインタラクションが可能になる．これにより施設の利用や商品の購入などの際に現実空間に近い体験を得ることが可能となる．そして，これらは観光分野でのユースケースとして考えられるもので，さらに教育・研修や視察といった領域へも応用可能であると考えている．

3.3 問題設定

モバイル環境での利用を前提にすると，それに伴う様々な問題が出現する．例えば，ネットワークの帯域や計算リソースが限られる場合がある．限られたリソースの中で QoE を最大化するため，品質に影響を与える要因を調整しなければならない．

本稿では，対象オブジェクトのスキャンに深度カメラを用いるが，得られるすべての点群をリアルタイムに送信すると利用可能な帯域を圧迫してしまう．例えば，Azure Kinect のカラー点群ストリームは，720P/15FPS で約 430MB/s

の帯域幅を使用する．3.1 節で述べたようにセルラー通信でリアルタイムに送信することはできない. 帯域の圧迫は，FPS の低下やその他データの遅延・不達を引き起こす可能性がある他，通信量の増大につながるためモバイルでの運用を想定する環境では許容できない．また点群データは点の数が多くなれば，その分だけ後の処理時間に影響を与え，リアルタイム性の低下に繋がる恐れがある．そのため，OOI（Object Of Interest）に応じて選択した点群データを送信する場合でも，ユーザーの QoE に大きな影響を与えずにデータを削減することが理想的である．

本稿では，4G/5G ネットワークやモバイルデバイスの利用といった条件下において，ユーザーの QoE に大きな影響を与えずにデータを削減し，点群のリアルタイムストリーミングを実現するという問題を解決する．

4. 提案システム

本章では，3 章で述べた要件を満たすシステムの設計と，その構成要素であるハードウェアや機能といった事項について述べる．

4.1 アプローチ

観光のオンライン化を実現するにあたり，現実空間（観光地）をリアルタイムにデータ化したうえで，仮想空間上で再構成する必要があるが，現実空間上のすべてのオブジェクトをリアルタイムに取得して利用することは不可能である．そのため Lee らのアプローチ [4] と同様，現実空間から一部分を切り取り，必要とされるオブジェクトのみをリアルタイムに反映させ，その他のオブジェクトについては静的なデータを使用するというアプローチを採用した．つまり現実空間を遠隔参加者が投影を希望するオブジェクトとその背景となる周囲のオブジェクトに分割した．**図 1** にこのアプローチを示す．なお，このときのオブジェクトはそれぞれ OOI(Object Of Interest)，AO(Ambient Object) とする．

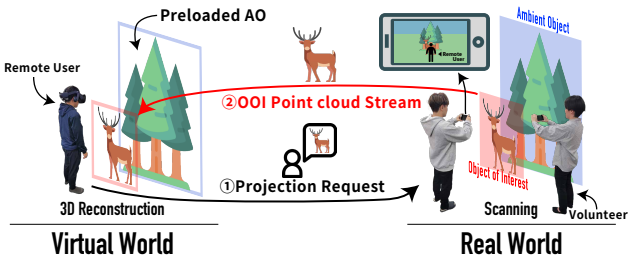

図 1 提案システムのコンセプト

また 3.1 節で示したように．OOI 点群のスキャンや送信は．ボランティアの協力のもと，スマートフォンやタブレット端末などのモバイル端末で行うことを想定している．このような条件下では，OOI 点群のみを送信する場合でも，通信帯域や計算資源の観点から，QoE を低下させる遅延が発生する可能性がある．そこで，OOI 点群の品質を動的に制御し，使用リソースを実用可能な範囲に抑える手法を提案する．従来のタイリング手法 [5], [7] では，点群オブジェクトを均等に分割していたが，提案手法では対象物（人間や動物）の身体・骨格部位の変動性を考慮して点群の品質（つまり解像度やフレームレート）を動的に制御する．オブジェクトの属性を認識し，体の部位単位で点群を分割することで，オブジェクトのきめ細かい品質制御を行い，帯域消費を抑えながらユーザ体験を最大化することができる．

4.2 ハードウェア構成

提案システムでは，オブジェクトのリアルタイムスキャンにスマートフォンやタブレットといったモバイルデバイスの使用を想定しているが，本稿では，点群ストリーミングの性能を評価するため，深度カメラ：Azure Kinect も使用した．この Azure Kinect は Jetson AGX Xavier に接続され，ROS を介したデータの受け渡しや，OOI の 3 次元切り出し等の処理を行う．また，この Jetson とは別に Windows PC を用意し，互いに LAN で接続した．この Windows PC は，提案システムにおけるリモート側端末を模倣しており，予め用意した AO と，Jetson から送られる OOI を重ねて表示するために構成した．**図 2** にシステム構成を示す．なお AO のモデル作成には LiDAR 搭載の iOS デバイスを使用している．この詳細については 4.4 項にて述べる．

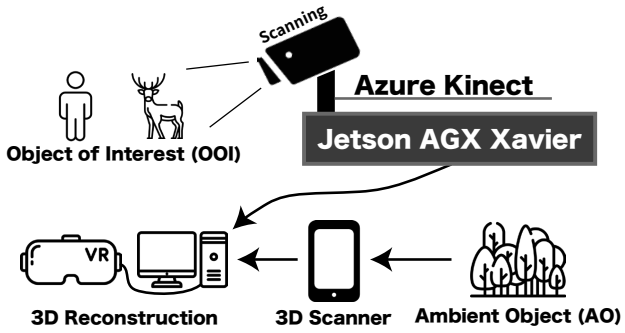

図 2 ハードウェア構成図（品質評価用）

4.3 システムデザイン

実空間のオブジェクトをスキャン・転送し，仮想空間にリアルタイムに再構成することで遠隔地とローカルユーザーが空間を共有するシステムを提案する．システムの概要を図 1 に示す．観光地にいるボランティアが遠隔ユーザの要請に応じて現実空間のリアルタイムスキャンを行う．ボランティアが現地のオブジェクト（OOI）に対してスキャニングとデータの送信をリアルタイムに行い，予め用意され

た背景データ (AO) と併せて遠隔ユーザの仮想空間内に現実空間を再構成することで，観光地の空間を共有する．

また，システムで扱う点群データ (OOI) の帯域幅消費を減らすために，身体部位の変動性を考慮した動的な品質制御を取り入れる．このアプローチは 2D 映像における H.265 のような動画圧縮規格のアイデアと 3D 点群におけるタイリングアプローチを参考にするもので，オブジェクトが大きく動いているときは高フレームレート/低解像度とし，逆にオブジェクトが静止に近い状態にあるときは低フレームレート/高解像度にするというように，動作に応じて各部位の解像度とフレームレートを動的に変化させることで点群データの総量を削減するものである．これら一連の処理により，点群データ (OOI) は生データよりもデータ量が小さくなり，絶対的な品質としては低下するが，ユーザの QoE にできるだけ影響を与えないように品質制御を行うことで，帯域幅消費の最適化を行う．

4.4 実装

我々が上記システムの実現可能性について評価するために作成した試作システムについて説明する．**図 3** にシステム概要を示す．なお本システムは ROS 上に構築され，また 3D 再構成には Unity を使用している．

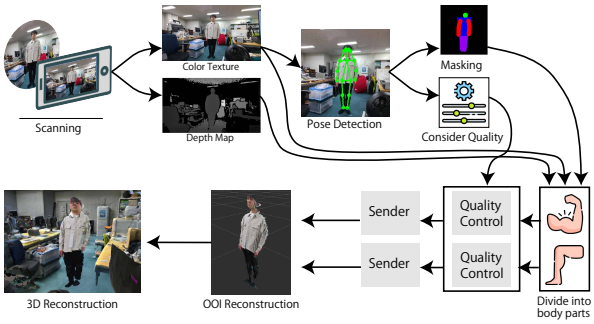

図 3 System Data Flow

まず仮想環境内で現実空間を再構成する際に使用する AO と OOI について，その取得方法を説明する．まず再構成時に背景として使用される AO であるが，iPhone に搭載される LiDAR を活用して 3D スキャンを行うアプリケーションを利用したり，360 度画像を活用するなど複数の手法により取得可能である．なお近年，様々な地点で 3D 化が進められているため，将来的には AO オブジェクトのスキャンは必要なくなる可能性がある．**図 4** に AO の取得例を示す．

次に，実際に伝送及びリアルタイム再構成の対象となる OOI 点群の取得方法について述べる．使用デバイスとしては，前述の通りモバイルデバイスによる取得を想定している．Apple が iPhone や iPad に LiDAR センサーをリリースしたことで，3D スキャニングが多くの人々にとって容易

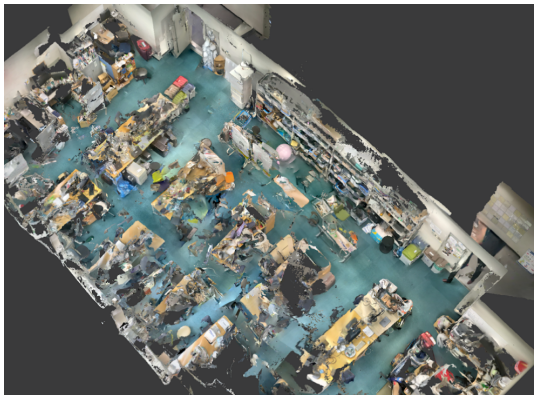

図 4 点群例（環境対象）

なものとなった．デプスカメラを使用して OOI の深度とテクスチャを連続的に取得しカラー点群として扱う．OOI オブジェクトのスキャン後，そのコンテキストに応じて動的に行う品質制御について説明する．まず図 3 に示すように，カメラから入力される 2D カラー映像に対して骨格推定を行い，身体部位の区分けと移動量の追跡を行う．なお骨格推定について，人間の骨格推定には trt-pose[8], [9] の学習済みモデルを使用し，動物（鹿）の骨格推定では**図 5** に示すように DeepLabCut[10], [11] にて学習したモデルを使用している．点群の品質は，対応する身体部位の変位の

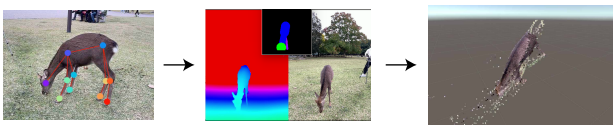

図 5 動物の場合（鹿）：動物の姿勢推定結果からマスクを生成し点群を得る

値によって決定される．システムは常に現在のフレームとその前のフレームの骨格位置の差異を監視している．この差異が閾値を上回った時にシステムは身体部位が動いてると判断して即座に該当部位の解像度を下げてフレームレートを優先する．逆に，身体部位の移動量が閾値を下回る状態が数フレーム続いたときは，オブジェクトが静止に近いと判断し，段階的に解像度を上げていく（フレームレートが下がる）．品質の種類に関しては，オリジナルの入力（カラー画像・デプスマップ：720P）に対して最近傍補完を用いたリサイズ（25%，50%，75%の 3 段階）を行うことで点群の解像度を調節し，同時に送信タイミングを調整することでフレームレートも切り替えている．なおこの処理を各身体部位ごとに行う．**表 1** では，サンプルとして用意した人間の映像シーケンスに対して，我々のシステムが品質を制御した際の帯域幅消費を示す．このサンプルでは，体の全部位を 75%の品質としたときの帯域幅消費（つまり最大値）は，20.2Mbps となり，既存のインフラでも十分送信可能な範囲内である．

これにより得られた各身体部位の点群データは ROS ト

表 1　提案システムにより出力される点群品質の仕様

Compression Rate (Original:100%)	Head	Body	Arms	Legs	All
25%	0.28Mbps (13FPS)	2Mbps (13FPS)	2.3Mbps (14FPS)	1.5Mbps (14FPS)	6.1Mbps
50%	0.9Mbps (12FPS)	5.5Mbps (9FPS)	6.1Mbps (9FPS)	4.8Mbps (12FPS)	17.3Mbps
75%	1.8Mbps (9FPS)	6.5Mbps (5FPS)	6.4Mbps (5FPS)	5.6Mbps (6.5FPS)	20.2Mbps

ピックとして Unity へ送信され，予め用意された AO と併せて 3D 再構成を行う．また，点群を表現する際に水密性を確保するため，点群の品質によって点の大きさを動的に変更している．このようにして再構成された観光地の空間は，VR-HMD を通してユーザに提示される．なお 3D 再構成では OOI と AO の座標を一致させる必要があるが，本研究の範囲外である．評価システム内では手動で座標の一致を行った．この問題は点群のマッチング/レジストレーションや AR 情報を組み合わせることで解決ができると考えている．

5.　評価実験

この章では，品質制御がユーザの QoE に与える影響を評価するために行った実験について示す．**図 6** に実験の概要を示す．

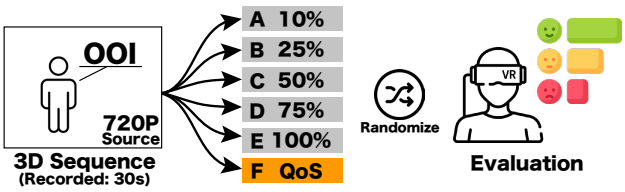

図 6　品質評価実験

5.1　実験概要

この実験では，我々が予め用意した 30 秒ほどの点群シーケンスについて，複数の異なる品質レベルの出力をユーザに見てもらい，各品質に対する評価を得る．なお，この点群シーケンスは，人間が立っている（静止している），手を振っている，足踏みをしているの 3 状態で構成されている．

また上記の点群映像に加え，骨格部位の変動性を考慮して動的に品質を切り替えた場合の点群映像も評価の対象とし，合計 6 パターンの点群映像評価とした．（条件 F：提案システム）

図 7 に実験シーンを示す．本実験では VR ヘッドセット（HTC-VIVE）を用いて，3 次元空間内のコンテンツを表示した．

5.1.1　評価映像

各圧縮率に対する点群の例を**図 8** に示す．なおここでの圧縮率は，システムに入力する映像に対するリサイズの倍率である．最近傍補間を用いてカラー・デプスの両映像のリサイズを行った．（圧縮率 0.1 の場合は 720P の 10%，つまり 128 × 72 の品質となる．）

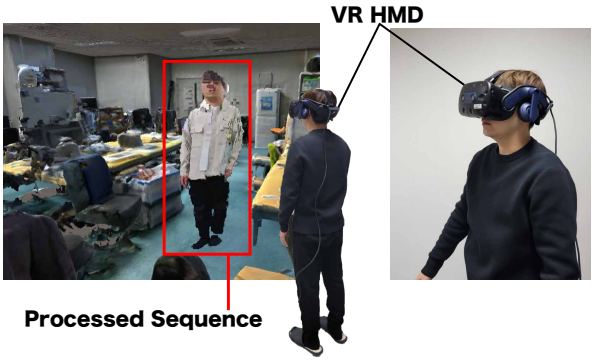

図 7　実験のイメージ図

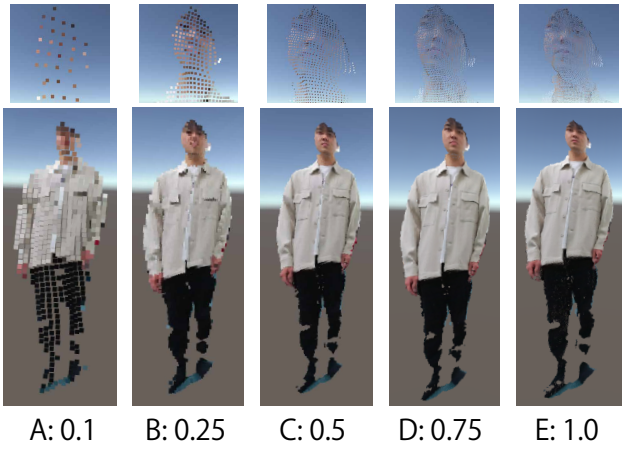

図 8　圧縮率の違いによる OOI 点群の変化

5.1.2　評価方法

評価手法は，代表的な映像品質評価手法の一つである ACR (Absolute Category Rating) 法 [12] を参考にしている．実験参加者は，10 秒程度の評価映像を観視した後，続く 10 秒以内に 5 段階品質尺度（5. 非常に良い，4. 良い，3. 普通，2. 悪い，1. 非常に悪い）により評価を行う．映像品質評価値がその前に観視した映像の品質に影響を受ける順序効果があるため，評価映像の提示順序をランダマイズして実験参加者毎に変化させている．評価内容は，(i) 映像の滑らかさに対する品質，(ii) 点群の緻密さに対する品質，(iii) 総合的な映像品質の 3 つである．

実験参加者は，20 代男性の大学生院生 12 名である．各実験参加者は，6 つの評価映像をランダムに視聴し，各映像を視聴した直後に 3 つの観点について 5 段階で評価をしている．

6.　結果

各評価項目に対する品質評価結果を**図 9**, **図 10**, **図 11** に示す．実験参加者から得られた評価結果をもとに，MOS を算出し可視化した．まず，各評価項目の結果について，

図 9 Quality for Smoothness

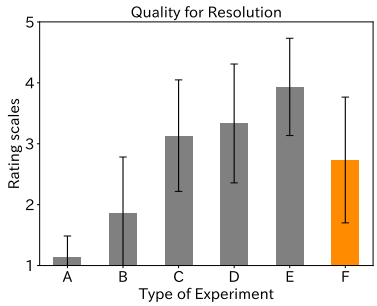

図 10 Quality for Resolution

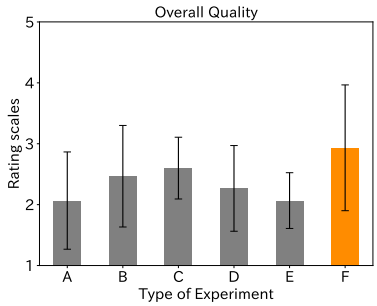

図 11 Overall Quality

Kruskal-Wallis 検定を用いた．検定の結果．3 つの実験とも滑らかさ・解像度・総合的な品質に有意な差が見られた．(p 値 Smoothness: 8.94×10^{-12}, Resolution: 1.14×10^{-9}, Overall Quality: 1.45×10^{-2}) その後，Mann-Whitney U-test を用いて，各基準結果の品質を比較した．検定結果を**表 2** に示す。滑らかさの品質結果に対する U 検定の結果，

表 2 Significant differences

	Two pairs to compare	P-value	Significance
Smoothness	Compression A,F	0.008	✓
	Compression B,F	0.790	
	Compression C,F	0.002	✓
	Compression D,F	0.000	✓
	Compression E,F	0.000	✓
Resolution	Compression A,F	0.001	✓
	Compression B,F	0.007	✓
	Compression C,F	0.244	
	Compression D,F	0.020	✓
	Compression E,F	0.001	✓
Overall Quality	Compression A,F	0.046	✓
	Compression B,F	0.138	
	Compression C,F	0.129	
	Compression D,F	0.026	✓
	Compression E,F	0.012	✓

提案手法である F は，B（圧縮率 0.25）との有意差が認められなかった．このことは，提案手法が条件 B と同程度の滑らかさを保持していることを示している．解像度の品質に対する U 検定の結果，提案手法である F は，C（圧縮率 0.5）との有意差が認められなかった．このことは，提案手法が条件 C と同程度の解像度を保持していることを示している．総合品質の品質結果に対する U 検定の結果，提案手法である F は，B（圧縮率 0.25）と C（圧縮率 0.5）との有意差が認められなかった．このことは，提案手法が条件 B と条件 C 同程度の総合品質を保持していることを示している．ここで，総合品質で同程度だった条件 B と C について詳しく検討すると，条件 B は，滑らかさは同程度だが，解像度は提案手法の評価が高い．条件 C は，解像度は同程度だが，滑らかさは提案手法の評価が高い．これらのことから，総合評価では同程度の評価になったものの，個別評価では提案手法が優れていることが確認された．

7. 議論

この評価実験では，3 次元点群を VR-HMD で視聴し品質の評価を行った．しかし，点群品質の主観評価は参加者の視覚能力やコンテンツへの好みなど，人間心理や視聴条件によって複雑になる．この実験でも ITU-T 勧告 [12] を参考に視覚能力の確認や映像順序のランダム化などを行ったが，この複雑さの軽減には限界がある [13]．特に個人の好みに関しては，鮮明な映像を好むグループと滑らかな映像を好むグループに大別されることが明らかになっている [14]．そのため我々の研究でもより厳密に QoE を評価するためには被験者やコンテンツによって品質評価を調整する必要があると考えられる．また今回の評価手法である ACR 法は一重刺激であり，二重刺激と比較すると評価のばらつきが大きくなりがちである．しかし，VR 上で二重刺激法を採用するのは困難である．これは同時に同じ条件で評価映像を提示することができないためで [15], [16]，3D コンテンツの QoE 評価は今後も検討を続ける必要がある．

8. まとめ

本論文では観光の仮想化を実現するために，人や動物といった現実空間のオブジェクトを対象に，仮想空間上でリアルタイムに 3D 再構成を行い，現実空間（観光地）を遠隔地へ 3 次元で空間共有するための点群ストリーミング手法を提案した．現実空間の人間や動物といったオブジェクトを 3D 点群としてデータ化し，モバイル環境での限られたネットワーク帯域や計算リソースの条件下でシステムを実現するため，身体部位のうち激しく動く部分は解像度を下げてフレームレートを上げる/動かない部分は解像度を上げてフレームレートを下げる，といったように骨格部位の変動性を考慮して点群の解像度やフレームレートを調整する動的品質制御のアプローチを提案し，実装と評価を行った．その結果，システムは品質制御により，事前に用意した点群シーケンスを最大 20.2Mbps で転送可能となり，既存のセルラーネットワークでの点群ストリーミング（アップロード）が可能であることを示した．また評価実験の結果，提案手法を動的品質制御の無い場合と比較したとき，

総合評価では同程度の評価になったものの，滑らかさや解像度といった個別評価では提案手法が優れていることが確認された．

謝辞　本研究の成果は NICT 課題 222「ウイルス等感染症対策に資する情報通信技術の研究開発」の支援によるものです．

参考文献

[1] 石丸大稀, 中村優吾, 藤本まなと, 諏訪博彦, 安本慶一ほか. 実空間の人や動物とのリアルな遠隔コミュニケーションのための qos-aware 点群ストリーミング手法. マルチメディア, 分散, 協調とモバイルシンポジウム 2022 論文集, Vol. 2022, pp. 1527–1533, 2022.

[2] NTT ドコモ. 実効速度計測結果 — 通信・エリア — ntt ドコモ. `https://www.docomo.ne.jp/area/effective_speed/`.

[3] Alaeddin Nassani, Li Zhang, Huidong Bai, and Mark Billinghurst. Showmearound: Giving virtual tours using live 360 video. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–4, 2021.

[4] Yongjae Lee, Byounghyun Yoo, and Soo-Hong Lee. Sharing ambient objects using real-time point cloud streaming in web-based xr remote collaboration. In *The 26th International Conference on 3D Web Technology*, pp. 1–9, 2021.

[5] Jie Li, Cong Zhang, Zhi Liu, Wei Sun, and Qiyue Li. Joint communication and computational resource allocation for qoe-driven point cloud video streaming. In *ICC 2020-2020 IEEE International Conference on Communications (ICC)*, pp. 1–6. IEEE, 2020.

[6] Zhi Liu, Susumu Ishihara, Ying Cui, Yusheng Ji, and Yoshiaki Tanaka. Jet: Joint source and channel coding for error resilient virtual reality video wireless transmission. *Signal Processing*, Vol. 147, pp. 154–162, 2018.

[7] Jounsup Park, Philip A Chou, and Jenq-Neng Hwang. Rate-utility optimized streaming of volumetric media for augmented reality. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, Vol. 9, No. 1, pp. 149–162, 2019.

[8] Alexander Mathis, Pranav Mamidanna, Kevin M. Cury, Taiga Abe, Venkatesh N. Murthy, Mackenzie W. Mathis, and Matthias Bethge. Deeplabcut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, 2018.

[9] Tanmay Nath*, Alexander Mathis*, An Chi Chen, Amir Patel, Matthias Bethge, and Mackenzie W Mathis. Using deeplabcut for 3d markerless pose estimation across species and behaviors. *Nature Protocols*, 2019.

[10] Bin Xiao, Haiping Wu, and Yichen Wei. Simple baselines for human pose estimation and tracking. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 466–481, 2018.

[11] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7291–7299, 2017.

[12] P ITU-T RECOMMENDATION. Subjective video quality assessment methods for multimedia applications. 1999.

[13] Shishir Subramanyam, Jie Li, Irene Viola, and Pablo Cesar. Comparing the quality of highly realistic digital humans in 3dof and 6dof: A volumetric video case study. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 127–136. IEEE, 2020.

[14] Prem Nandhini Satgunam, Russell L. Woods, P. Matthew Bronstad, and Eli Peli. Factors affecting enhanced video quality preferences. *IEEE Transactions on Image Processing*, Vol. 22, No. 12, pp. 5146–5157, 2013.

[15] Evangelos Alexiou, Irene Viola, Tomás M Borges, Tiago A Fonseca, Ricardo L De Queiroz, and Touradj Ebrahimi. A comprehensive study of the rate-distortion performance in mpeg point cloud compression. *APSIPA Transactions on Signal and Information Processing*, Vol. 8, , 2019.

[16] Irene Viola and Touradj Ebrahimi. A new framework for interactive quality assessment with application to light field coding. In *Proceedings of the SPIE, Volume 10396, id. 103961F 17 pp. (2017)*, p. 51, 09 2017.