連合強化学習を用いた点群映像ストリーミングに関する研究

研究代表者　　劉　志　　電気通信大学　准教授

1 はじめに

点群映像は、3D 空間上に表現された多数の点で構成されており、各点には座標や色などの複数の属性が関連付けられています。点群映像システムは、人間に 6 自由度（現在の最高レベル）の臨場感を提供する視聴体験を可能にし、次世代の映像技術として注目されています。本研究では、点群映像の通信および計算における制約を考慮し、連合強化学習に基づく点群映像のストリーミング手法を検討しました。実験の結果、提案手法が非常に優れた性能を示すことが確認されました。

2 点群映像ストリーミングの背景

2-1 点群映像

点群は 3 次元空間上の点で構成されており、各点には座標や色などの複数の属性が関連付けられています。点群データの分布に応じて、点群の符号化手法は大きく 2 つのカテゴリに分類されます［1, 2］。分布が比較的均一な点群は、既存の 2 次元映像技術を用いて 2D フレームに投影することが可能ですが、疎に分布する点群データは階層的なキューブ（立方体）に分割され、各点がその所属するキューブのインデックスとして符号化されます。なお、点群の符号化処理は従来の映像に比べて計算コストが高いという特徴があります。点群は、最も有望なボリュメトリック映像フォーマットの一つとされており、学術界および産業界の両方から大きな注目を集めています［2,3］。点群映像を配信するためには、帯域幅に応じた画質適応、効率的な画質評価指標、そして正確な 6 自由度（6DoF）視野（Field of View: FoV）予測などが求められます［4, 5］。例えば、点群映像を毎秒 30 フレームでストリーミングする場合、生の点群映像のサイズが非常に大きいため、必要な帯域幅は最大で 6Gbps にも達することがあり、効率的な伝送手法が強く求められます［6］。

2-2 点群映像の品質評価

点群映像の品質評価は、最適化の目的関数を定義するものであり、点群映像伝送システムにおける重要な構成要素です。従来の映像伝送においては、ピーク信号対雑音比（PSNR）が一般的に用いられており、これは基準フレームと復元フレームの画素ごとの差異を評価することで画質を定量化する指標です。また、Video Multimethod Assessment Fusion（VMAF）は、主観的評価を反映した高度な指標として広く利用されています［7, 8］。点群映像の品質評価においても、これらと類似のロジックに基づく指標がいくつか提案されています。たとえば、文献［9］では、MPEG PCC［10］の品質評価ソフトウェアを用いて、点と点の歪みに基づく PSNR を算出し、客観的な品質評価を行っています。また、文献［11］では、表現の基礎的な品質、ユーザーの視点からの詳細度、デバイスの解像度などに基づく効用（utility）指標を導入しています。さらに、文献［12］では、拡張現実（AR）環境における動的点群に対して、ITU-T P.1203 モデル［13］を調整した評価モデルを提案しています。しかしながら、これらの手法の多くは基本的に従来の映像品質評価指標を踏襲しており、点群映像特有の特徴を十分に考慮したものではありません。

2-3 点群映像ストリーミング

近年、点群映像ストリーミングは注目を集めており、さまざまな研究が行われています［4, 14］。これらの研究の多くは、仮想現実（VR）映像のストリーミング手法［15, 16, 17］に類似したアプローチを採用しており、映像を複数のタイルに分割し、ユーザーの FoV 内にあるタイルのみを送信することで、定義された目的関数の最適化を図っています［11, 18］。文献［14］では、点群映像の伝送最適化において、エンコーディング処理の高い計算負荷を考慮しています。しかしながら、これらの手法はいずれもモデルベースであり、動的なネットワーク状況への適応性には欠けているという課題があります。

2-4 グアダプティブビットレートストリーミング

近年、アダプティブビットレート（ABR）ストリーミングにおいて、強化学習（Reinforcement Learning: RL）の活用が注目されています [19, 20]。M. Claeys ら [21] は、環境との相互作用を通じてユーザーの QoE（Quality of Experience）を最適化する、強化学習ベースの HTTP アダプティブストリーミングクライアントを設計しました。また、360 度映像のストリーミングにおける RL ベースのレート適応も、優れた性能を示しています [22]。

この分野における主な課題としては、報酬関数の適切な定義や学習アルゴリズムの選定が挙げられます。さらに、従来のローカル学習手法では、トレーニングデータセットの規模が個々のユーザーデバイスに限定されるという重大な制約があります。通信環境はセッションごとに変化するため、各デバイスが十分なデータを収集し、様々なネットワーク状況に対応できるよう学習するには長い時間が必要となります。 Pensieve [19] が指摘しているように、再学習（リトレーニング）の頻度は、モデルの汎化性能や新たなネットワーク挙動の出現頻度に依存します。Oboe [23] は、異なるネットワーク環境において Pensieve の性能を維持するためには再学習が不可欠であることをさらに示しました。

2-5 プライバシーの懸念

ユーザープライバシーの保護は、これまでの研究ではほとんど考慮されていません。トレーニングに必要なトラフィックメタデータ（例えば、伝送速度など）は、ユーザーのプライバシーを深刻に侵害する可能性があります。例えば、攻撃者はトラフィック速度の変化からユーザーの活動を推測することができます[24]。さらに、ユーザーの視線データからは、性別、年齢、認知状態、さらには精神的な障害といった大量の個人情報が明らかになる可能性があります [25]。したがって、分散されたユーザーデータセットを収集する従来の中央クラウドベースのトレーニング手法は、ユーザーのプライバシーを危険にさらすことになります。

3 連合強化学習を用いた点群映像ストリーミング

3-1 概要

私たちは FRAS(Federated reinforcement learning empowered adaptive point cloud video streaming) と呼ばれる方法を提案します. FRAS は、点群ストリーミングを適応的に配信するビデオ・オン・デマンドシステムであり、ユーザーにより高い QoE（Quality of Experience：体感品質）を提供します。図 3 に示すように、点群映像は有線ネットワークを通じてサーバに送信され、タイル分割されます。その後、分割されたタイルは複数の品質レベルにサンプリングされ、圧縮されます。サーバは、これらすべてのタイル（圧縮・非圧縮の両方）を用いて、各映像に対する MPD（Media Presentation Description）ファイルを生成し、クライアントがオンデマンドで映像を再生する際に HTTP 経由で提供します。各クライアントは、ユーザーの 6 自由度（6DoF）の姿勢、ネットワーク状況、バッファの状態、および MPD ファイルから取得したタイル情報などの環境情報に基づいて、強化学習（RL）ベースのレート適応手法により、必要なタイル ID および品質レベルを選択します。その後、クライアントはネットワークインターフェースを通じて選択したタイルを取得します。

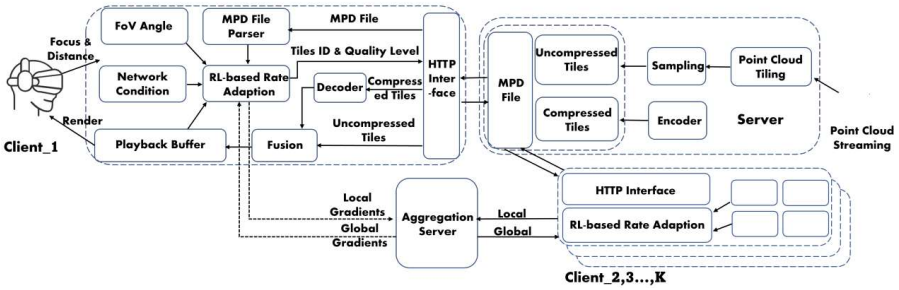


図 1：システム構成。

一方、学習フェーズにおいては、複数のクライアントが FRL（Federated Reinforcement Learning）アルゴリズムを通じて勾配情報を集約サーバに共有します。なお、映像ソースとサーバ間の帯域幅は一般的に十分にあるため、本研究ではサーバとクライアント間の通信に焦点を当てます。この通信には、クライアント i における時刻 t における制限的かつ動的な帯域幅 B_{i,t}が関与します。ここで、t は、複数のフレームが 1 つのグループとしてエンコードされる基本時間単位であり、これは従来の動画ストリーミングにおける GOP（Group of Pictures）と同様です。帯域幅 B_{i,t}は、セクション 3-3 で説明した帯域予測手法により推定されます。

クライアント i は、時刻 t において FoV F_{i,t}を持ち、ユーザーが要求する最低品質 Q_i を下回らない映像を求めています。ユーザー i は、時刻 t において、圧縮タイルをデコードするための利用可能な計算資源 C_{i,t}を持っています。クライアントは、6DoF の FoV、帯域幅、再生バッファ、およびデコード処理の複雑さに基づいて、事前に定義された報酬ポリシーに従って RL を活用し、QoE の最適化を図ります。

3-2 タイル分割とダウンサンプリング

クライアントの視野（FoV）内でのスムーズな視野切り替えと高品質な映像表示を実現するために、点群映像は均質なタイルに分割されます。計算および通信効率を考慮し、FoV 内のタイルは FRL アルゴリズムによって選択されたエンコードレートで送信し、FoV 外のタイルは最低品質でエンコード・送信されます。まず、対象となる点群空間を最適に囲む直方体を認識し、その 3 次元サイズを特定します。その後、高さ方向に垂直な平面上で N×M の分割を行い、高さ方向に H 層に分割します。最終的に、全体で N×M×H のタイルが得られます。各時刻 t におけるこれらの分割タイルは、さらに均等にダウンサンプリングされ、L 段階の品質レベルに分類されます。

3-3 FoV およびネットワーク予測

360 度映像では視点の自由度が 3 つ（ヨー、ピッチ、ロール）であるのに対し、点群映像では位置に関する 3 つの自由度（x, y, z）が加わるため、合計 6 自由度となります。これによりユーザーの没入感は向上しますが、FoV（視野）の予測がより複雑になります。本研究では、ユーザーの FoV を予測するために、6 自由度を個別に予測し、それらを統合するというシンプルかつ効果的な手法を採用しています。各次元の予測には、以下の 3 つの手法を試験しました：前回の FoV を使用する手法，線形回帰を用いる手法，多層パーセプトロン（Multilayer perceptron：MLP）を用いる手法。

点群映像のストリーミングにおいては、ネットワーク帯域幅の予測も重要です。本研究では、帯域幅予測の手法として、ゲート付きリカレントユニット（Gated Recurrent Unit：GRU）[26] と指数加重移動平均（Exponentially Weighted Moving Average：EWMA）[27] の 2 つを比較しています。GRU は、時系列予測に広く用いられる人工リカレントニューラルネットワークのアーキテクチャであり、ロング・ショートターム・メモリ（Long short-term memory： LSTM）に似ていますが、フォゲットゲート（忘却ゲート）を備えており、LSTM と比較して計算複雑度が低く、わずかに高い性能を実現できる場合があります。一方、EWMA は時系列モデリングにおいてシンプルかつ強力な統計手法です。

3-4 制約条件

前述の通り、FRAS は FoV 外のタイルを除き、選択的な品質レベルで全てのタイルを取得します。FoV 外のタイルについては、最低品質で送信されるか、全く送信されない場合もあります。本システムには、以下の 3 つの制約条件があります：

- 各タイルは、圧縮された状態または非圧縮の状態で送信されなければなりません。
- 任意の時刻 t において、送信されるすべてのタイルの総データ量は、ネットワーク帯域幅を超えてはなりません。
- 送信されたすべてのタイルに対して必要なデコードリソースは、システムのデコードリソース上限（キャパシティ）を超えてはなりません。

3-5 ユーザースタディによる QoE モデリング

MPEG の点群圧縮標準 [38] に基づき、本研究ではジオメトリ歪みの指標として点対点 PSNR を採用しています。したがって、品質評価指標として主に PSNR、エンコード品質レベル、シーンからの距離を考慮しま

す。クライアントがシーンに近づくにつれて、ユーザーは各タイルの PSNR やその品質レベルに対してより敏感になります。したがって、品質構成要素の重みは距離に応じて変化すべきです。よって、時刻 t における クライアント i の FoV（視野）を F_{i,t}とすると、その受信映像の品質は以下のように定義されます：これは、FoV 内の各タイルについて、ユーザーの距離を考慮しながら、エンコード品質レベル l と PSNR q の加重平均を示すものです。

4 FRL による適応的タイル品質選択

本セクションでは、通信および計算リソースの制約下で QoE（ユーザー体感品質）を最大化するために、適切なタイル品質レベルを適応的に選択する手法として採用したフェデレーテッド強化学習(FRL:Federated Reinforcement Learning）について説明します。

最適化目標には、既存研究と同様にリバッファリング時間 T_r や品質の滑らかさ（品質変動）Δ_t も考慮します。さらに、本研究では点群映像のデコードの複雑性にも注目しています。従来の映像ストリーミングでは、デコード遅延は非常に小さく、リバッファリング時間 T_r に含める必要はありませんでした。しかし、モバイルデバイスでの映像ストリーミングに関する近年の研究では、QoE モデルにおいてコーディング効率を考慮する動きが始まっています。本研究では、従来の映像ストリーミングと比較して、点群映像のストリーミングにおいてデコード遅延が特に重要であると強調します。その理由は以下の 2 点です：点群映像タイルのデコード処理は、従来の映像に比べてはるかに高い複雑性を持つことが、我々のデコードテストにより証明されています。

前述の通り、点群映像ストリーミングでは、デコード負荷を軽減するために圧縮タイルと非圧縮タイルの選択を取り入れています。そのため、デコード遅延を QoE モデルに組み込むことは、この問題をより効果的に最適化する上で有効です。ただし、QoE にはすでに品質要素と時間要素が含まれているため、さらに第三の要素（デコード容量など）を追加するとモデルの複雑性および学習の難易度が大幅に上がるという問題があります。このため、本研究では、デコードに必要な容量を直接使用するのではなく、「デコード時間ペナルティ」T_d を QoE に導入します。これは、実際のデコード時間からチャンク（動画の単位）持続時間を差し引いた値を指します。複数の品質レベルを持つ様々な点群映像に対して多数のデコードテストを実施し、その平均値を本アルゴリズムに利用しています。これにより、点群映像ストリーミングにおける QoE は、以下のように定式化されます：

$$QoE_{dis} = f(Q, T_r, \Delta l, T_d)$$
$$= \alpha \times \sum_{FoV} q_{n,m,h,l,t} + \beta \times l - \gamma \times T_r - \delta \times \Delta l - \epsilon \times T_d,$$

4-1 ローカルトレーニングのための強化学習

各クライアントの QoE を最適化するために、セクション 4 で定義した目的関数に従って、強化学習（RL）アルゴリズムを用いてローカルモデルをトレーニングします。前述の通り、各クライアントは短時間のうちに限られたネットワーク環境に直面するため、汎化性と堅牢性のあるモデルを得るには、長期間にわたるデータ収集と学習が必要です。しかし、これは QoE の低下やデバイスのバッテリー寿命に悪影響を及ぼします。

そこで本研究では、フェデレーテッドラーニング（FL）と強化学習（RL）を統合し、ユーザー間の学習経験を集約することで、プライバシーを保ちつつ、データ拡張と学習の高速化を図ります。主な構成要素は以下の通りです：

- 状態（State）：強化学習エージェントの入力は、s_t = (l_{t-1}, b_t, B_t, d_t, n_t, r_t) です。各要素は以下の意味を持ちます：
  - l_{t-1}：前回の点群映像チャンクの選択されたエンコード品質レベル（非圧縮タイルには圧縮タイルと異なるレベルインデックスが割り当てられる）
  - b_t：現在のバッファレベル
  - B_t：予測されたネットワーク帯域幅

- d_t：現在のチャンクのダウンロード時間
- n_t：次の映像チャンクのサイズ候補のベクトル
- r_t：残りの映像チャンク数

- 行動（Action）：エージェントは状態 s_t を受け取り、次の点群映像チャンクのエンコード品質レベルを選択します。

- 学習手法：Actor-Critic アルゴリズムを用いてポリシーをローカルで学習します。報酬は、セクション 4 で定義された QoE に従って定義されます。累積割引報酬に関するポリシーの勾配は、式(7)により計算されます [28]。

4-2 FL アルゴリズム

本研究では、点群映像ストリーミングのための適応的なグループ化フェデレーテッドラーニング(Grouped Federated Learning) アルゴリズムを提案します。このアルゴリズムは、異種かつ動的なネットワーク環境に対応しつつ、ユーザープライバシーを保護することを目的としています。詳細は文献 [29] を参照してください。

5 シミュレーション

5-1 シミュレーション設定

（1） ビデオソース

トレーニングには、Panoptic Studio から提供された 36 本の点群ビデオシーケンスを使用しました。テストには、vsenseVVDB2 データベース 2 から提供された AxeGuy、LubnaFriends、Rafa2、Matis、Loot、Longdress の 6 本の点群ビデオを用いました。各点群フレームは 3×3×4 のタイルに分割され、それぞれ 20%、40%、60%、80%、100%の割合でサンプリングされます。これは品質レベル 1 から 5 に対応しています。エンコードおよびデコードには、KD-tree を用いた Google Draco コーデック 3 を使用しています。標準化された MPEG の V-PCC エンコーダ 4 ではなく Draco を使用する理由は、性能が優れており、現行のハードウェアでリアルタイムデコードが可能なためです。

(2) ネットワーク条件

点群ビデオは圧縮後でも高ビットレートであるため、実験における現実的なスループット変動を再現するためには高帯域幅のトレースを用いる必要があります。そこで、Lumos5G [30]が提供する商用の mmWave 5G スループットデータセットの帯域幅トレースを再生してネットワークをシミュレートしました。このデータセットは歩行および車両移動の 2 つのモビリティモードにおける 118 本のトレースを含み、合計 18 時間以上のアクティブモニタリングをカバーしています。テスト中、アルゴリズムは 3.3 節で述べたように予測帯域幅を用いて次のビデオチャンクのエンコード品質を選択し、選択された品質レベルと予測帯域幅およびその他の状態指標を用いて報酬を計算します。

(3) 視野（FoV）

本研究では、ViVo [31]が提供するビューポート軌跡データセットを使用します。これは 4 つのボリュメトリックビデオにおけるスマートフォンユーザーやヘッドセットユーザーを含む 32 人のユーザーのビューポート軌跡を提供しており、位置情報（x, y, z）と方向情報（yaw, pitch, roll）を含み、合計 1 万点以上のデータ項目があります。3.3 節に述べたように、各次元の予測には以下の 3 つの手法を用います。

1. LAST（Last FoV）：前の時点の FoV を現在の FoV として使用
2. LR（線形回帰）および

3. MLP（多層パーセプトロン）：過去 8 個のビューポートの履歴ウィンドウから現在の FoV を予測。MLP には 8 個および 2 個のニューロンを持つ 2 層の軽量ニューラルネットワークを用い、活性化関数は ReLU、重み最適化はLimited-memory BFGS(L-BFGS)[32]を使用しています。LR およびMLP はいずれも scikit-learn ライブラリで実装しています。

(4) QoE 評価

提案した QoE モデルの評価のために、既存の 2 つの QoE モデル（PSNR および POSTER）と比較しました。PSNR は画像品質評価で広く使われる客観的指標です。POSTER[33] はユーザーの主観的な品質により近い指標であり、視覚的品質やユーザー指向の指標を推定に利用し、VMAF[34]よりも優れた性能を示しています。QoE モデルの視聴距離固定の制約を補うため、中間距離（1.3m、1.7m、2.3m、2.7m）での値を補間により推定しました。

(5) モデルおよび評価指標

FRAS モデルでは、過去 12 回分の帯域幅測定値をサイズ 4、ストライド 1 の 128 個のフィルターを持つ 1 次元畳み込み層に入力します。次のチャンクのサイズ候補は同様の形状の別の 1D-CNN に入力されます。これらの層の出力は 128 ニューロンの隠れ層で統合され、ソフトマックス関数が適用されます。クリティックネットワークは同一構造ですが、出力層は線形ニューロンです。割引率はγ=0.99、アクターとクリティックの学習率はそれぞれ 10^-4 と 10^-3、エントロピー係数βは 5 から 0.1 へ 3×10^5 イテレーションにかけて減衰します。10 クライアントによる協調学習を行い、FRL アルゴリズムでグローバル勾配を集約する 1 台の集約サーバを設置しています。本構成は TensorFlow 2.3.0 で実装しました。

(6) ベースライン

性能比較のため、以下の 5 つのベースラインを実装しました。

・ ViVo[31]：6DoF の特性を考慮したモバイルボリュメトリックビデオストリーミングの包括的研究で、トリーミングを最適化。

・ QUETRA[35]：選択されたビットレート、ネットワークスループット、バッファ容量を用いた待ち行列モデルに基づき、期待バッファ占有率を計算するレート適応アルゴリズム。

・ Pensieve[19]を適用した強化学習（RL）手法：単一 RL モデルを訓練。

・ robustMPC[36]：予測スループットの誤差を正規化により考慮し、MPC を改善。

・ Buffer-Based（BB）[37]：起動フェーズのバッファ占有率に基づき品質レベルを選択。占有率が 0.1 秒を下回らないようにし、1 秒を超えた場合は最高品質を自動選択。

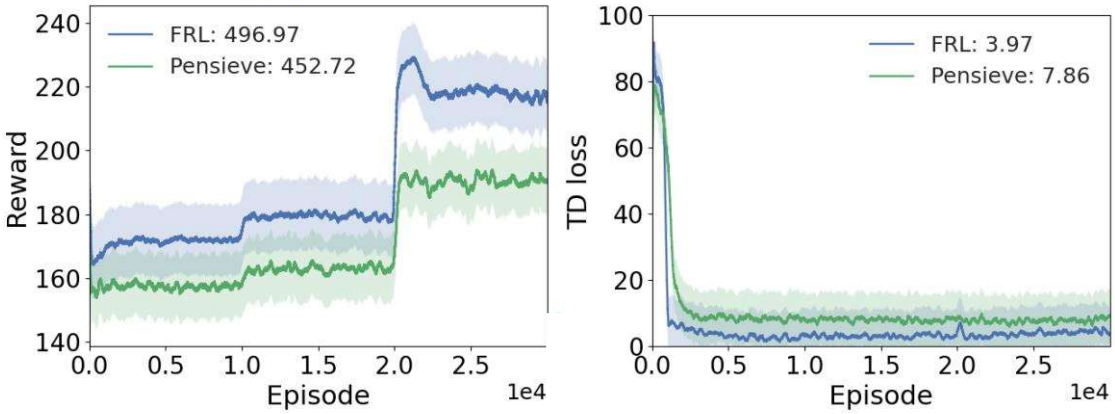
(7) 実装環境

上記ベースラインをサポートするため、PCL 1.9.1 と QT 5.12 を用いて点群ビジュアライザを構築しました。ビジュアライザは、各アルゴリズムを実装した ABR クライアントからの品質選択決定を取得するよう設定しています。プレーヤーは再生バッファ容量 5000ms に設定しています。各シーケンスは 300 フレーム、10 秒間の動画で、より現実的な再生シナリオを再現するためにループ再生します。また、実験では GOP を 10 に設定し、1 チャンクあたり約 330ms とし、短時間に多数のリクエストが送信されることを回避しています。ビデオプレーヤーと ABR サーバはクライアントと同一サーバ上で動作し、Intel Xeon Gold 6246R CPU、NVIDIA GeForce GTX 3090 GPU、480GB SSD を搭載しています。HTTP サーバは Python HTTP モジュールで構築し、タイル化された点群動画および対応する MPD ファイルを格納しています。実際のストリーミングシステムでは、ユーザーが方向を素早く変えられ、モーションシックネスを防止するため、全点群に対して最低品質の圧縮ベースレイヤを実装しています。

5-2 シミュレーション結果

（1） トレーニング

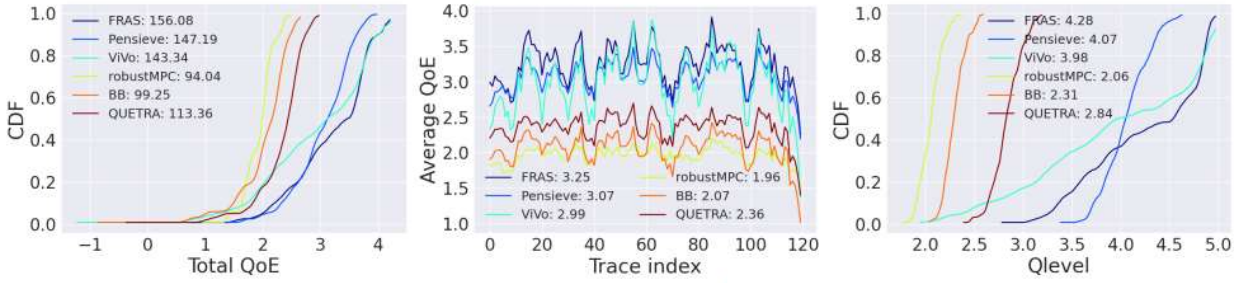
まず、コアアルゴリズムである FRL のトレーニング性能を、Pensieve から適応した RL アルゴリズムと比較して検証しました。図 2a および図 2b には、両アルゴリズムのトレーニング中の報酬および損失の推移を示しています。FRL の曲線はグローバルなフェデレーテッドラーニング（FL）モデルの値を表し、一方で RL の曲線は個別クライアントモデルの値を示しています。収束後の FRL は、報酬で 9.77%、損失で 49.50%上回り、Pensieve を凌駕しました。



(a) Reward　　(b) Loss

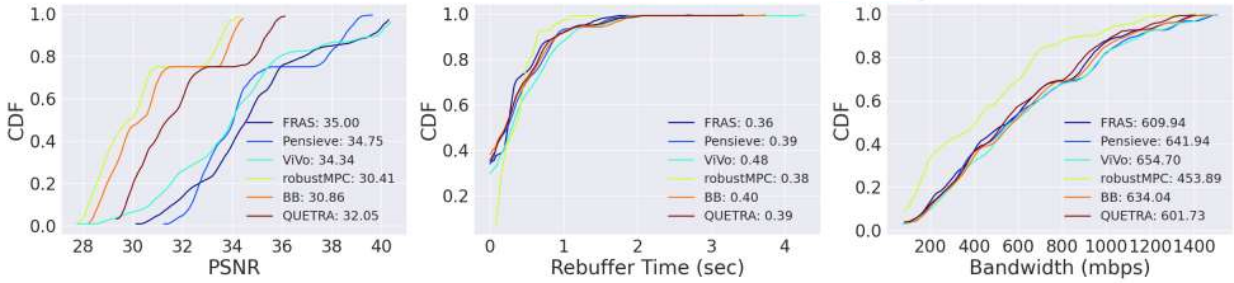
図 2：FRL と Pensieve の学習時の比較

（2） テスト



(a) Sum of QoE　　(b) Average QoE　　(c) Sum of quality level

図 3：QoE（ユーザ体験品質）のパフォーマンス



(a) PSNR　　(b) Rebuffering　　(c) Bandwidth

図 4：主要な指標のパフォーマンス

異なるネットワーク条件下で収集した帯域幅トレースを用いて、FRAS の性能をベースラインと比較して検証しました。図 3 および図 4 に示すように、FRAS はほとんどのシナリオにおいてすべての指標でベースラインを大幅に上回る結果を示しました。具体的には、FRAS は平均 QoE（報酬）において、robustMPC、BB、QUETRA、ViVo、Pensieve と比較してそれぞれ 66%、57%、38%、9%、6%の改善を達成しました。さらに、平均ビデオ品質レベルは robustMPC、BB、QUETRA、ViVo、Pensieve と比較してそれぞれ 108%、85%、51%、8%、5%向上し、平均 PSNR についても同様に 15%、13%、9%、2%、1%の改善を示しました。また、平均リバッファリング時間は QUETRA、robustMPC、BB、ViVo と比較してそれぞれ 67%、56%、33%、11%減少しました。

なお、全体で 5 段階のエンコード品質レベル（各レベルは圧縮版と非圧縮版が存在）を用いているため、ベースラインに対する改善は平均品質レベルの向上とほぼ一致し、PSNR の改善結果とも整合しています。このため、FRAS はベースラインよりも帯域幅の使用率が高くなっていますが、一方で全テストアルゴリズム中で最小のリバッファリングを示しており、これは FRAS が帯域幅の飽和頻度を最も低く抑えていることを示しています。したがって、FRAS は利用可能な帯域幅を最適に活用し、性能と QoE の最良のトレードオフを実現できることが証明されました。

6. 結論

本研究では、点群映像の通信および計算に関する制約を考慮し、連合強化学習に基づく点群映像ストリーミング手法について検討しました。実験により、提案手法が非常に良好な性能を示すことが確認されました。

なお、紙幅の都合により、本報告では主に上述の FRAS に関する内容に焦点を当てていますが、それ以外にも以下の複数の研究を行いました。

a. 点群映像における視点予測に関して：新たな映像顕著性検出手法を提案しました。この手法では、領域映像顕著性情報とユーザーの視聴履歴を用いた、視点予測を行っています。実験により、本手法は既存の最先端手法を上回る性能を達成したことが確認されました。

b. 映像伝送最適化に関して：従来手法では有限時間を前提とした最適化が一般的でしたが、本研究では無限時間領域を考慮した伝送最適化問題を定式化し、それに対応する最適化アルゴリズムを設計しました。シミュレーションの結果、提案手法は既存の手法に比べて大幅な性能向上を示しました。

【参考文献】

[1].Chao Cao, Marius Preda, and Titus Zaharia. 2019. 3D Point Cloud Compression: A Survey. In The 24th International Conference on 3D Web Technology. ACM, 1–9

[2].Sebastian Schwarz, Marius Preda, Vittorio Baroncini, Madhukar Budagavi, Pablo Cesar, Philip A Chou, Robert A Cohen, Maja Krivokuća, Sébastien Lasserre, Zhu Li, et al. 2018. Emerging MPEG standards for point cloud compression. IEEE Journal on Emerging and Selected Topics in Circuits and Systems 9, 1 (2018), 133–148

[3].Evangelos Alexiou, Evgeniy Upenik, and Touradj Ebrahimi. 2017. Towards subjective quality assessment of point cloud imaging in augmented reality. In 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP). IEEE, 1–6

[4].Alexander Clemm, Maria Torres Vega, Hemanth Kumar Ravuri, Tim Wauters, and Filip De Turck. 2020. Toward Truly Immersive Holographic-Type Communication: Challenges and Solutions. IEEE Communications Magazine 58, 1 (2020), 93–99

[5].Feng Qian, Bo Han, Jarrell Pair, and Vijay Gopalakrishnan. 2019. Toward Practical Volumetric Video Streaming on Commodity Smartphones. In Proceedings of the 20th International Workshop on Mobile Computing Systems and Applications. ACM, 135–140

[6].E d'Eon, Bob Harrison, T Myers, and PA Chou. 2017. 8i voxelized full bodies-a voxelized point cloud dataset. ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document WG11M40059/WG1M74006, Geneva (2017)

[7].Tianchi Huang, Chao Zhou, Rui-Xiao Zhang, Chenglei Wu, Xin Yao, and Lifeng Sun. 2019. Comyco: Quality-aware adaptive video streaming via imitation learning. In Proceedings of the 27th ACM international conference on multimedia. 429–437
[8].Netflix. 2019. VMAF—Video Multi-Method Assessment Fusion. https://github.com/Netflix/vmaf. Accessed: Sep. 20, 2019.
[9].Mohammad Hosseini and Christian Timmerer. 2018. Dynamic adaptive point cloud streaming. In Proceedings of the 23rd Packet Video Workshop. ACM, 25–30.
[10].Rufael Mekuria and Lazar Bivolarsky. 2016. Overview of the MPEG activity on point cloud compression. In 2016 Data Compression Conference (DCC). IEEE, 620–620
[11].Jounsup Park, Philip A Chou, and Jenq-Neng Hwang. 2019. Rate-utility optimized streaming of volumetric media for augmented reality. IEEE Journal on Emerging and Selected Topics in Circuits and Systems 9, 1 (2019), 149–162.
[12].Minh Nguyen, Shivi Vats, and Hermann Hellwagner. 2024. No-Reference Quality of Experience Model for Dynamic Point Clouds in Augmented Reality. In Proceedings of the 3rd Mile-High Video Conference. 90–91
[13].R ITUTP. 2017. Parametric bitstream-based quality assessment of progressive download and adaptive audiovisual streaming services over reliable transport-quality integration module. International Telecommunication Union (2017)
[14].Jie Li, Cong Zhang, Zhi Liu, Wei Sun, and Qiyue Li. 2020. Joint Communication and Computational Resource Allocation for QoE-driven Point Cloud Video Streaming. IEEE International Conference on Communications
[15].Ching-Ling Fan, Shou-Cheng Yen, Chun-Ying Huang, and Cheng-Hsin Hsu. 2021. On the Optimal Encoding Ladder of Tiled 360° Videos for Head-Mounted Virtual Reality. IEEE Transactions on Circuits and Systems for Video Technology 31, 4 (2021), 1632–1647.
[16].Pantelis Maniotis and Nikolaos Thomos. 2021. Tile-Based Edge Caching for 360° Live Video Streaming. IEEE Transactions on Circuits and Systems for Video Technology 31, 12 (2021), 4938–4950.
[17].Michael Zink, Ramesh Sitaraman, and Klara Nahrstedt. 2019. Scalable 360° Video Stream Delivery: Challenges, Solutions, and Opportunities. Proc. IEEE (2019)
[18].Jeroen van der Hooft, Tim Wauters, Filip De Turck, Christian Timmerer, and Hermann Hellwagner. 2019. Towards 6dof http adaptive streaming through point cloud compression. In Proceedings of the 27th ACM International Conference on Multimedia. 2405–2413
[19].Hongzi Mao, Ravi Netravali, and Mohammad Alizadeh. 2017. Neural adaptive video streaming with pensieve. In Proceedings of the Conference of the ACM Special Interest Group on Data Communication. 197–210
[20].Wenqian Xue, Jialu Fan, Victor G Lopez, Jinna Li, Yi Jiang, Tianyou Chai, and Frank L Lewis. 2019. New methods for optimal operational control of industrial processes using reinforcement learning on two time-scales. IEEE Transactions on Industrial Informatics (2019)
[21].Maxim Claeys, Steven Latre, Jeroen Famaey, and Filip De Turck. 2014. Design and evaluation of a self-learning HTTP adaptive video streaming client. IEEE communications letters 18, 4 (2014), 716–719.
[22].Nuowen Kan, Junni Zou, Chenglin Li, Wenrui Dai, and Hongkai Xiong. 2022. RAPT360: Reinforcement Learning-Based Rate Adaptation for 360-Degree Video Streaming With Adaptive Prediction and Tiling. IEEE Transactions on Circuits and Systems for Video Technology 32, 3 (2022), 1607–1623
[23].Zahaib Akhtar, Yun Seong Nam, Ramesh Govindan, Sanjay Rao, Jessica Chen, Ethan Katz-Bassett, Bruno Ribeiro, Jibin Zhan, and Hui Zhang. 2018. Oboe: auto-tuning video ABR algorithms to network conditions. In SIGCOMM

[24].Noah Apthorpe, Dillon Reisman, Srikanth Sundaresan, Arvind Narayanan, and Nick Feamster. 2017. Spying on the smart home: Privacy attacks and defenses on encrypted iot traffic. arXiv preprint arXiv:1708.05044 (2017)
[25].Myrthe Faber, Robert Bixler, and Sidney K D'Mello. 2018. An automated behavioral measure of mind wandering during computerized reading. Behavior Research Methods 50, 1 (2018), 134–150
[26].Rui Fu, Zuo Zhang, and Li Li. 2016. Using LSTM and GRU neural network methods for traffic flow prediction. In 2016 31st Youth Academic Annual Conference of Chinese Association of Automation (YAC). IEEE, 324–328
[27].Pantelis Maniotis and Nikolaos Thomos. 2021. Tile-Based Edge Caching for 360° Live Video Streaming. IEEE Transactions on Circuits and Systems for Video Technology 31, 12 (2021), 4938–4950.
[28].Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. 2016. Asynchronous methods for deep reinforcement learning. In International conference on machine learning. PMLR, 1928–1937
[29].Gao, Y., Zhou, P., Liu, Z., Han, B., & Hui, P. (2022). Fras: Federated reinforcement learning empowered adaptive point cloud video streaming. arXiv preprint arXiv:2207.07394.
[30].Arvind Narayanan, Eman Ramadan, Rishabh Mehta, Xinyue Hu, Qingxu Liu, Rostand A. K. Fezeu, Udhaya Kumar Dayalan, Saurabh Verma, Peiqi Ji, Tao Li, Feng Qian, and Zhi-Li Zhang. 2020. Lumos5G: Mapping and Predicting Commercial MmWave 5G Throughput. In Proceedings of the ACM Internet Measurement Conference (Virtual Event, USA) (IMC '20). Association for Computing Machinery, New York, NY, USA, 176–193
[31].Bo Han, Yu Liu, and Feng Qian. 2020. ViVo: Visibility-aware mobile volumetric video streaming. In Proceedings of the 26th annual international conference on mobile computing and networking. 1–13
[32].Dong C Liu and Jorge Nocedal. 1989. On the limited memory BFGS method for large scale optimization. Mathematical programming 45, 1 (1989), 503–528
[33].Wangyu Choi and Jongwon Yoon. 2024. Poster: User-Oriented QoE Model for Video Streaming on Mobile Deivces. In Proceedings of the 22nd Annual International Conference on Mobile Systems, Applications and Services. 662–663
[34]. Netflix. 2019. VMAF—Video Multi-Method Assessment Fusion. https://github.com/Netflix/vmaf. Accessed: Sep. 20, 2019
[35].Praveen Kumar Yadav, Arash Shafiei, and Wei Tsang Ooi. 2017. Quetra: A queuing theory approach to dash rate adaptation. In Proceedings of the 25th ACM international conference on Multimedia. 1130–1138
[36].Xiaoqi Yin, Abhishek Jindal, Vyas Sekar, and Bruno Sinopoli. 2015. A control-theoretic approach for dynamic adaptive video streaming over HTTP. In Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication. 325–338
[37].Te-Yuan Huang, Ramesh Johari, Nick McKeown, Matthew Trunnell, and Mark Watson. 2014. A buffer-based approach to rate adaptation: Evidence from a large video streaming service. In Proceedings of the 2014 ACM conference on SIGCOMM. 187–198

〈発　表　資　料〉

題　　名	掲載誌・学会名等	発表年月
VPFormer: Leveraging Transformer with Voxel Integration for Viewport Prediction	ACM Transactions on Multimedia Computing,	2025.04

in Volumetric Video	Communications, and Applications	
Spatial Quality Oriented Rate Control for Volumetric Video Streaming via Deep Reinforcement Learning	IEEE Transactions on Circuits and Systems for Video Technology	2024.12
Optimal Multibitrate Video Caching and Processing in Edge Computing: A Stackelberg Game Approach	IEEE Internet of Things Journal	2025.07
Horizon: a Real-time Point Cloud Video Streaming System over Wireless Networks	21st ACM International Conference on Mobile Systems, Applications, and Services (ACM MobiSy)	2023.06
Land-scape: Saliency and Trajectory based Viewport Prediction in Point Cloud Video Streaming	21st ACM International Conference on Mobile Systems, Applications, and Services (ACM MobiSy)	2023.06