Virtual Reality(VR)ビデオネットワーク伝送に関する研究

研究代表者　　劉　志　　静岡大学学術院工学領域数理システム工学系列　助教

1 はじめに

仮想現実(VR)ビデオは人間に臨場感がある視聴体験を提供するので、エンターテインメント、教育、医療などの分野でのVRアプリケーションが多数登場している。VR ビデオストリーミングとは、 VR ビデオコンテンツをネットワークに通して転送、再生するダウンロード方式の一種である。 VRビデオストリーミングは、多数の VRアプリケーションを実現するために不可欠である。VRビデオコンテンツの巨大なサイズ、ビュー(view, ユーザが見る部分) 切り替えのゼロ遅延の要件と無線ネットワークパケットロスがあるために、ユーザにVR ビデオストリーミングを提供するのは非常に困難である。

本研究では、既存無線ネットワーク上でユーザに高品質の VRサービスを提供するために、 アプリケーション層、物理層などをクロスレイヤ最適化する VR ビデオストリーミング方式を提案する。次に、OFDMAシステムにおける360 VR映像(i.e.ビデオ)最適マルチキャストとプロトタイプシステムについて説明する。

2 OFDMA システムにおける 360 VR 映像の最適マルチキャスト

2-1 背景

全方位カメラを用いて、興味のあるシーンを全方位同時に撮影することで、VR映像を生成することができる。撮影した映像を3D球体にステッチしてワープさせ、投影法を用いて2Dマップに投影する。このようにして得られた映像を 360VR(or VR)映像と呼ぶ。最も使用されている投影方法は、3D 球体を矩形に投影する(i.e. equirectangular projection) である［1］。VR ヘッドセットやヘッドマウントディスプレイ (HMD)を装着したユーザは、いつでも自由に任意の方向から見たいシーンを見ることができ、没入感のある視聴体験を楽しむことができる。VRはエンターテインメント、教育、医療などの分野での応用が広い、2020年にはVR市場は300億ドルに達すると予測されている［2］。

360VRビデオストリーミングは、多数のVRアプリケーションを実現するために不可欠である。360VR映像は、従来の映像に比べてサイズが大きいため、360VR 映像を丸ごと送信することは、無線ネットワークに大きな負担をかける。また、360VR映像を視聴しているユーザーは、どの時点でも、1つ視聴する角度を選択し、全体のビデオの 1 つの部分(i.e. Field of view (FoV))を見る。そのため、360VR映像全体を送信することは不要である。そこで、360VR 映像の伝送効率を向上させるために、タイリング(tiling)技術が広く採用されている。具体的には、360VR 映像を同じ大きさの長方形のセグメントに分割し、これをタイル(tile)と呼び。予測された将来の各 FoV をカバーするタイルのセットを送信することで、ユーザの QoE(quality of experience) を低下させることなく、通信リソースを節約することができる。VRユーザーは、視聴方向を変更することができる。予測誤差を許容するために、より大きなタイルセットを送信する。

本研究では, 360 VR ビデオのワイヤレス伝送に注力している。これまでの研究[3], [4]では、TDMA システムおよび OFDMA システムにおいて、マルチキャストの機会を利用して、それぞれ単一品質の360 VR 映像を最適伝送することを検討した。[5]-[7]では, シンガーユーザ無線ネットワークにおける多品質360VR映像の伝送を検討し, 伝送するタイルごとに最適な品質レベルを選択することに着目している. [5]-[7]で提案された解決策では, 不均一なチャネル条件を持つユーザ間での最適なリソース共有が考慮されておらず, マルチキャストの機会が無視されているため, システムのパフォーマンスが低下する可能性がある。[8]、[9]では、複数のユーザに向けて多品質360VR映像を送信するシナリオを検討し、マルチキャストの機会を利用して送信効率を向上させ。具体的には、[8]では、全ユーザの総効用が最大となるように、送信するタイル毎に品質レベルの選択を最適化し。この最適化問題は、タイルを個別に考慮するため、そのサイズが不必要に大

公益財団法人電気通信普及財団
研究調査助成報告書　第35号　2020年度

きくなる。我々の先行研究[9]では、全ユーザの総効用を最大化するための最適品質レベル選択を検討していたが、タイルをグループ化よる計算複雑度を効果的に低減することができる。しかし，[8]，[9]で提案されている単一タイムスケール解は，実用的なシステムでは望ましい性能が得られない可能性がある．また、[8]，[9]は自然なマルチキャストの機会のみを利用しており、トランスコーディング（FFmpeg などのトランスコーディングツールを用いて、ある品質レベルのタイルをより低い品質レベルのタイルに変換すること）は考慮していません。ユーザ側のトランスコーディングはマルチキャストの機会を創出し、通信リソースを節約することができる。一方、ユーザ側でのトランスコーディングは計算リソースを消費する。システム全体のリソースを節約するために、トランスコーディングある場合のマルチキャスト機会を最適に創出する方法は未解決の問題として残っている。

本研究では、ユーザーごとにトランスコードを行う場合と行わない場合で、複数のユーザーにマルチクオリティの 360VR 映像マルチキャストを検討する。

2-2 提案手法概要

ユーザ側でのトランスコーディングにより、より多くのマルチキャストの機会が生まれ、システム性能を向上させることができる．これに基づいて、 ユーザ側でのトランスコーディングを行う場合と行わない場合を検討し、TDMA システムをモデル化し、最適化問題に定式化する。

(1) トランスコーディングなしでの最適なトランスミッション

トランスコーディング行わない場合、以下のように問題を定式化する。

Problem 1 (Energy Minimization without User Transcoding):

$$
\begin{aligned}
E^\star \triangleq \min_{\mathbf{t},\mathbf{p}} \quad & \mathbb{E}\left[\sum\nolimits_{\mathcal{S}\in\mathcal{I}}\sum\nolimits_{l\in\mathcal{L}} t_{\mathbf{H},\mathcal{S},l}\, p_{\mathbf{H},\mathcal{S},l}\right] \\
\text{s.t.} \quad & t_{\mathbf{h},\mathcal{S},l} \geq 0, \quad \mathbf{h}\in\mathcal{H}^K,\ \mathcal{S}\in\mathcal{I},\ l\in\mathcal{L}. \\
& \sum\nolimits_{\mathcal{S}\in\mathcal{I}}\sum\nolimits_{l\in\mathcal{L}} t_{\mathbf{h},\mathcal{S},l} \leq T, \quad \mathbf{h}\in\mathcal{H}^K. \\
& p_{\mathbf{h},\mathcal{S},l} \geq 0, \quad \mathbf{h}\in\mathcal{H}^K,\ \mathcal{S}\in\mathcal{I},\ l\in\mathcal{L}. \\
& |\mathcal{P}_{\mathcal{S}}| D_{r_k} \leq \frac{B}{T}\mathbb{E}\left[t_{\mathbf{H},\mathcal{S},r_k}\log_2\left(1+\frac{p_{\mathbf{H},\mathcal{S},r_k} H_k}{n_0}\right)\right], \\
& \qquad\qquad \mathcal{S}\in\mathcal{I},\ k\in\mathcal{S}.
\end{aligned}
$$

これを等価な凸問題に変換し、標準的な凸最適化手法を用いて最適解を求める。詳細は論文[10]を参照してください。

(2) ユーザートランスコーディングを用いた最適なトランスミッション

次に、トランスコーディングを行う場合を考える。ユーザトランスコーディングを用いてシステムを定式化し、定式化された問題は以下のようになる。

Problem 3 (Energy Minimization with User Transcoding):

$$
\begin{aligned}
\overline{E}^{\star} \triangleq \min_{\mathbf{y},\mathbf{t},\mathbf{p}} \quad & \mathbb{E}\left[\sum\nolimits_{\mathcal{S}\in\mathcal{I}}\sum\nolimits_{l\in\mathcal{L}} t_{\mathbf{H},\mathcal{S},l} p_{\mathbf{H},\mathcal{S},l}\right] \\
& + \beta \sum\nolimits_{\mathcal{S}\in\mathcal{I}}\sum\nolimits_{k\in\mathcal{S}} |\mathcal{P}_{\mathcal{S}}| E_k \left(\sum\nolimits_{l\in\mathcal{L}} l y_{\mathcal{S},k,l} - r_k\right) \\
\text{s.t.} \quad & t_{\mathbf{h},\mathcal{S},l} \geq 0, \quad \mathbf{h}\in\mathcal{H}^K,\ \mathcal{S}\in\mathcal{I},\ l\in\mathcal{L}. \\
& \sum\nolimits_{\mathcal{S}\in\mathcal{I}}\sum\nolimits_{l\in\mathcal{L}} t_{\mathbf{h},\mathcal{S},l} \leq T, \quad \mathbf{h}\in\mathcal{H}^K. \\
& p_{\mathbf{h},\mathcal{S},l} \geq 0, \quad \mathbf{h}\in\mathcal{H}^K,\ \mathcal{S}\in\mathcal{I},\ l\in\mathcal{L}. \\
& y_{\mathcal{S},k,l} \in \{0,1\}, \quad \mathcal{S}\in\mathcal{I},\ k\in\mathcal{S},\ l\in\mathcal{L}, \\
& \sum\nolimits_{l\in\mathcal{L}} y_{\mathcal{S},k,l} = 1, \quad \mathcal{S}\in\mathcal{I},\ k\in\mathcal{S}. \\
& \sum\nolimits_{l\in\mathcal{L}} l y_{\mathcal{S},k,l} \geq r_k, \quad \mathcal{S}\in\mathcal{I},\ k\in\mathcal{S}, \\
& |\mathcal{P}_{\mathcal{S}}| D_l y_{\mathcal{S},k,l} \leq \frac{B}{T}\mathbb{E}\left[t_{\mathbf{H},\mathcal{S},l}\log_2\left(1+\frac{p_{\mathbf{H},\mathcal{S},l}\acute{H}_k}{n_0}\right)\right], \\
& \qquad \mathcal{S}\in\mathcal{I},\ k\in\mathcal{S},\ l\in\mathcal{L}.
\end{aligned}
$$

この問題は、困難な discrete-continuous 最適化問題である。我々は、この問題を Difference of Convex (DC) programming に変換し、DC アルゴリズムを用いて最適解を求める。詳細は論文[10]を参照してください。

2-3 シミュレーション結果

ここでは、トランスコーディングを用いない場合の提案スキームを Proposed-w/o、トランスコーディングを用いた場合を Proposed-w とする。各提案スキームについて１つのベースラインスキームを選択し、ベースラインをそれぞれ Baseline-w/o, Baseline-w と表記する。 パラメータの詳細は論文[10]を参照してください.

(1) エネルギー消費量　vs ユーザ数

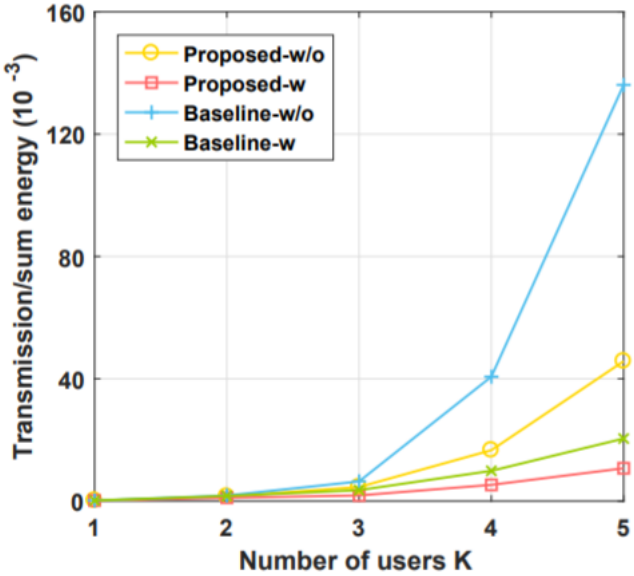

Figure 1: transmission energy vs user number

ここでは、エネルギー消費量とユーザー数の関係を図１に示している。利用者数に応じてトラフィック負荷が増加のため、各スキームのエネルギー消費量は、利用者数に応じて増加することがわかる。提案された方式は、最適化の利点により、ベースライン方式よりも優れている。さらに、トランスコーディングにより、マルチキャストの機会を増やし、システムの性能をさらに向上させ。

(2) エネルギー消費量　vs Zipf 指数

Zipf 指数が大きくなると、ユーザーの要求は同じになる傾向がある。送信エネルギーとユーザー数の関係

を図 2 に示している。ユーザーのリクエストが同じ傾向にある場合、トラフィックが少なくなり、消費電力が少なくなる。図 1 と同じ，提案された方式は、最適化の利点により、ベースライン方式よりも優れている。さらに、トランスコーディングにより、マルチキャストの機会を増やし、システムの性能をさらに向上させ。

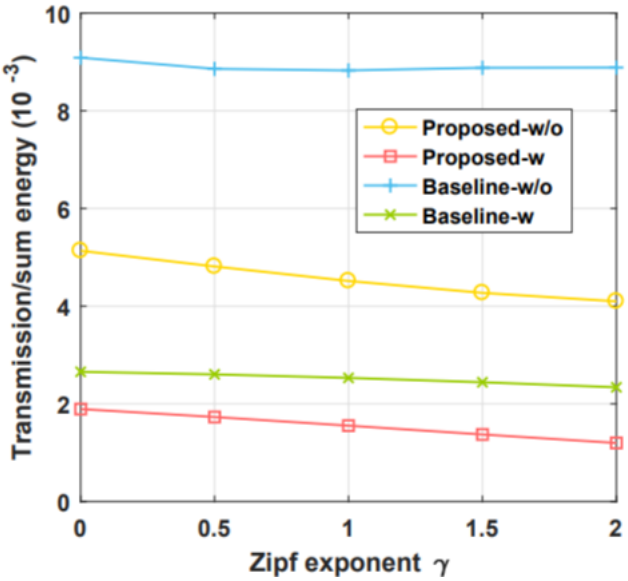

Figure 2: energy consumption vs Zipf exponent

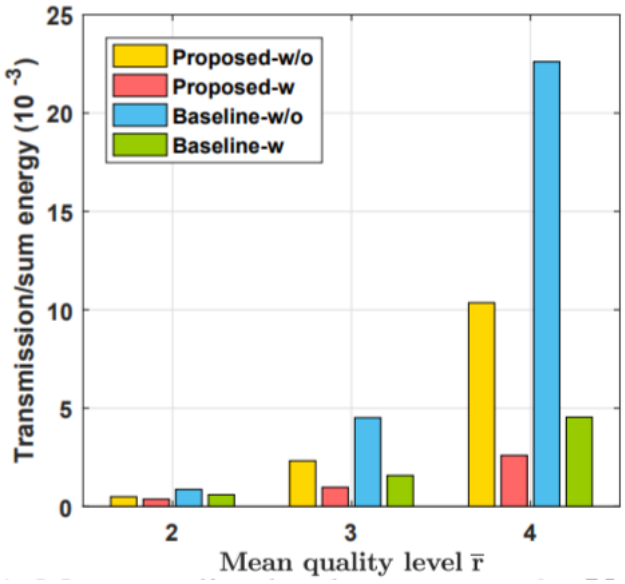

Figure 3: energy consumption vs mean quality

(3) エネルギー消費量　vs 平均品質レベル

エネルギー消費量と平均品質レベルの関係を図 3 に示している。要求された映像の平均品質レベルに応じてトラフィック負荷が増加のため、各スキームのエネルギーは、平均品質レベルに応じて増加することがわかる。提案された方式は、最適化の利点により、ベースライン方式よりも優れている。さらに、トランスコーディングにより、マルチキャストの機会を増やし、システムの性能をさらに向上させ。

3 プロトタイプシステム

3-1 プロトタイプシステム概要

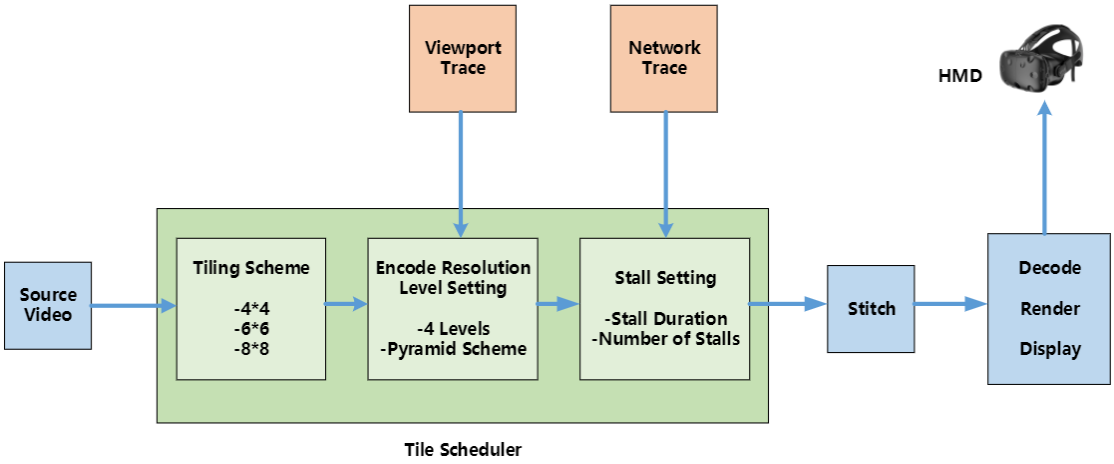

Figure 4: 360 VR video prototype system

- システムの概要: 図 4 に示すように、提案する 360VR 映像配信システムは、以下のように動作する。まず、サーバは、360VR 映像をタイルに分割し、各タイルを異なる品質にエンコードする。次に、サーバは、帯域幅予測とユーザビューポート予測の結果に基づいて、どのタイルを送信するか、および選択された各タイルの品質レベルを決定する。最後に、すべてのタイルがクライアントに到着すると、それらはデコードされ、レンダリングされ、ユーザに表示される。ユーザの QoE を向上させるために、サーバはユーザの FoV 内のタイルを可能な限り高画質で送信する。無線帯域幅が変動する場合、サーバは、それに応じて各タイルの品質レベルを調整する。具体的には、帯域幅が低下すると、サーバは、まず、ユーザのビューポート外のタイルの品質レベルを低下させ、次に、ビューポート内のタイルの品質レベルを低下させ、その後、品質切替現象が発生する[11]。帯域幅が全タイルの総ビットレート以下になるまで帯域幅が減少し続けると、バッファが空になり、その後に失速（stall）現象が発生する。
- ビューポート予測:ビューポート予測は、360VR 映像ストリーミングシステムの重要な部分です。360VR 映像を視聴する過程で異なるユーザのビューポート情報を取得し、そのビューポート情報に基づいてビューポート予測を行い。具体的には、予測の精度を向上させるために、360VR 映像実験者６０名（男 40 名、女 20 名）を選択し、各タイムインスタンスで 60 名ユーザのビューポート情報を取得する。同一のタイムインスタンスにおいて、異なるユーザの視野角が異なる場合がある。通常の状況では、異なるユーザのビューポートの重なりが発生し、最も重複確率の高い領域を中心に予測ビューポートを算出する。

Figure 5: Snapshot of the some original 360 videos used in thetes

- ネットワークトレース：本システムでは、LTE ネットワークトレース[12]を用いて、帯域予測モデ

ルを作成した（100%の確率で予測を行うことを前提とする）。また、本システムは、どのような帯域予測モデルにも適用可能である。

- 360VR 映像ソース：図 5 および表 1 に示すように、360 映像データベース［13］から 10 本の映像を選択する。

TABLE I: Parameters of ten original videos used in the test.

Sequence	Video Name	Resolution	Frame Rate	Duration
1	Cartoon	3840 × 2160	24fps	23s
2	Movie	3840 × 1920	30fps	25s
3	Surfing	3840 × 1920	25fps	23s
4	Soccer	3840 × 1920	25fps	20s
5	Basketball	3840 × 2048	30fps	25s
6	CS	3840 × 1920	30fps	25s
7	Aurora	4096 × 2048	12fps	22s
8	Car show	4096 × 2048	30fps	25s
9	Musical	3840 × 1920	30fps	20s
10	Gliding	3840 × 1920	25fps	25s

- 実験装置：各参加者はヘッドマウントディスプレイ（HMD）を装着して 360 度の映像を視聴する。実験では、HMD 装置として HTC VIVE を使用し、FoV は 110 度である。我々は、異なる方向から 360 のビデオを見るために 60 人の参加者のために回転可能な座席を提供する。
- 360VR 映像処理：図 5 に示すように、一般的に使用されている 360 映像データベース[13]から 10 本のソース映像を選択し、それぞれが高解像度で 30 秒から 40 秒の間の持続時間を持つ。これらの映像は、YouTube と VRCun からダウンロードしたものである。 本実験では、各 360 映像を 4*4, 6*6, 8*8 タイル（つまり、それぞれ 16、36、64 タイル）に分割する。各タイルは、解像度の観点から 5 段階にエンコードされている。タイリングと HEVC ビデオエンコーディングには GPAC と Kvazaar[14]を使用している。具体的なビデオタイリングスキームを図 6 および表 2 に示す。

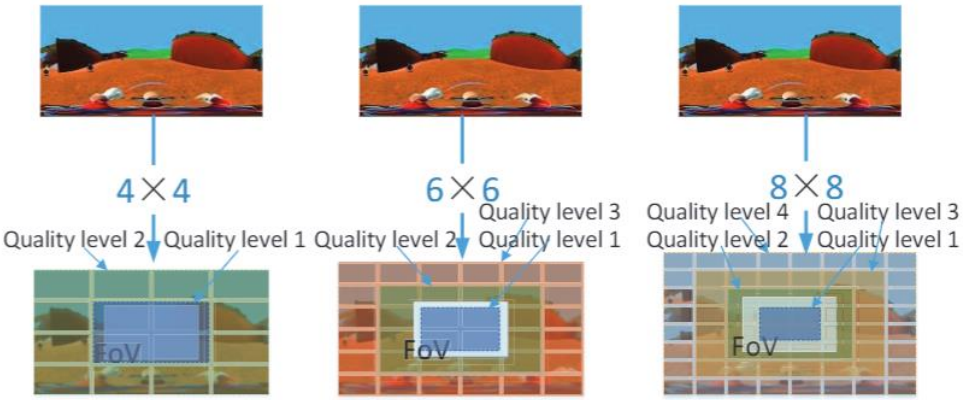

Figure 6:Tiling and pyramid resolution scheme.

TABLE II: Level of 360 video qualities. Level 1,2,3,4 have the same frame rate

Level	Resolution	Frame Rate
1	640 × 360	30fps
2	848 × 780	30fps
3	1280 × 720	30fps
4	3840 × 2160	30fps
5	4096 × 2048	30fps

3-2 QoE モデルと QoE 駆動のストリーミング最適化

本システムにより、VR 映像ストリーミングの QOE モデルを得ることができる。QoE は VR ビデオストリーミングの最適化のための品質指標として利用することができる。

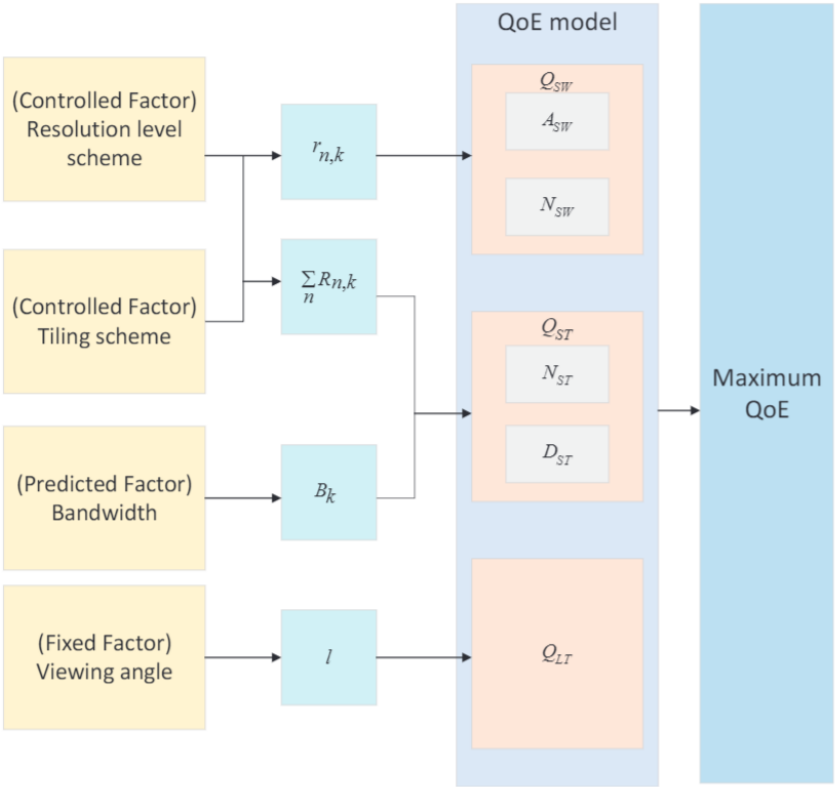

Figure 7: QoE driven 360 VR video streaming

提案された QoE メトリックは、緯度(latitude)、タイリング、失速、解像度レベルの切り替えなどの要因を含む。緯度はユーザの視野角によって制御される。タイリングは、ビデオサーバーのスティッチングや処理方法に影響される。失速や解像度レベルのスイッチは、無線ネットワークの帯域幅の影響を受ける.これを踏まえて、図 7 に示すように，予測可能な帯域幅と制御可能な映像処理方法を考慮した QoE 駆動の 360 映像ストリーミングフレームワークを提案する．このように、予測された帯域幅と固定された視野角が与えられると、QoE 値を最大化するための適切なタイリングと解像度レベルを見つけることができる。理論解析とシミュレーションを行い、性能を検証した。2 つのベースラインスキームを比較し、シミュレーション結果は以下の通りである。2 つのベースライン方式をそれぞれ Layered model[16]と Adaptive mode[17]と呼ぶ。

(1) QoE vs タイル数

QoE 値とタイル数 N*N の関係を調べ、その結果を図 8 に示す。シミュレーションでは、タイルの最大数を

1*1 から 6*6 まで変化させた。タイルを細かくすることで、FoV 内の映像コンテンツを伝送するための帯域幅を多く確保できる。一方、FoV 外のタイルは画質が悪くエンコードされているため、帯域幅の消費を抑えることができる。このようにして、ユーザは、より良い品質のビデオを得ることができ、より良い経験を楽しむことができる。

また、我々の手法は、ベースライン方式と比較して、より良い QoE の結果を得ることができる。

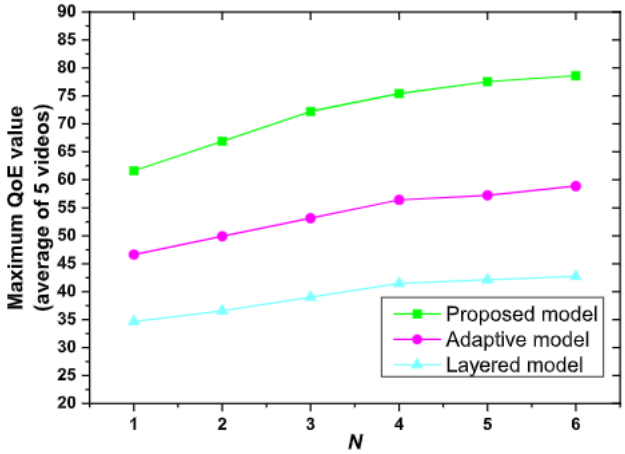

Figure 8: QoE vs number of tiles

(2) QoE vs 品質レベルの数

次に、利用可能な品質レベルの数を変えた場合の QoE 値を図 9 に示す。最適化された QoE 値は、品質レベルの数が増えるにつれて増加することがわかる。より多くの品質レベルが利用可能で、十分な無線帯域幅があれば、ユーザーはより良いビデオ品質を受信することができ、より良い QoE につながることは理にかなっている。また、我々の手法は、ベースライン方式と比較して、より良い QoE の結果を得ることができる。

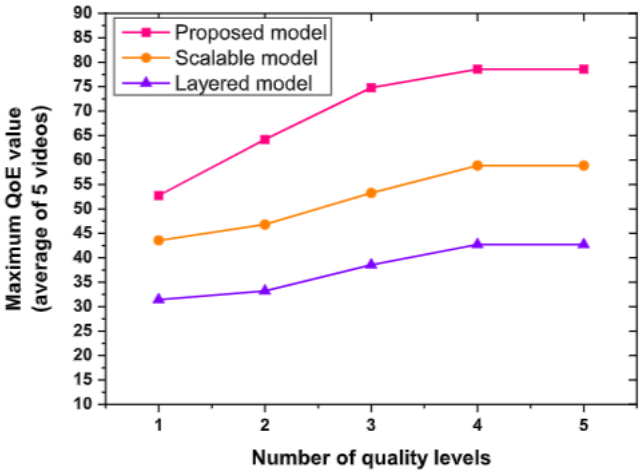

Figure 9: QoE vs quality levels.

4. まとめ

本研究では、既存無線ネットワーク上でユーザに高品質の VR サービスを提供するために、アプリケーション層、物理層などをクロスレイヤ最適化する VR ビデオストリーミング方式を提案した。

TDMA システムにおいて，ユーザ毎にトランスコードを行う場合と行わない場合の 360VR 映像を複数ユーザに最適伝送することを検討した．トランスコーディングを行わない場合には，自然なマルチキャストの機会を利用して平均伝送エネルギーが最小となるように伝送資源配分を最適化し，凸最適化手法を用いて最適解

を得る．また、ユーザがトランスコーディングを行う場合には、自然マルチキャストとトランスコーディング可能なマルチキャスト機会の両方を利用して、送信リソースの割り当てと送信品質レベルの選択を最適化し、DC Programming を用いて最適解を得る。

また、VR 映像システムを構築し、視野する角度、タイル数、品質レベルの切り替えを考慮した QoE モデルを得る。また、主観的品質評価実験を通じて、QoE とこれらの要因との関係を数学的に解析し、360 映像の総合的な QoE モデルを提示した。最後に、QoE を利用した 360 映像配信において、無線ネットワーク上においても映像品質の改善を確認した。

【参考文献】

[1] M. Zink, R. Sitaraman, and K. Nahrstedt, “Scalable 360 video stream delivery: Challenges, solutions, and opportunities,” Proc. IEEE, 2019.

[2] “Augmented virtual reality revenue forecast revised to hit 120 billion by 2020,” https://goo.gl/nw9mtP, 2016.

[3] C. Guo, Y. Cui, and Z. Liu, “Optimal multicast of tiled 360 vr video,” IEEE Wireless Commun. Lett., vol. 8, no. 1, pp. 145–148, Aug. 2019.

[4] C. Guo, Y. Cui, and Z. Liu “Optimal multicast of tiled 360 VR video in OFDMA systems,” Optimal multicast of tiled 360 VR video in OFDMA systems, vol. 22, no. 12, pp. 2563–2566, Dec. 2018.

[5] L. Xie, Z. Xu, Y. Ban, X. Zhang, and Z. Guo, “360probDASH: Improving QoE: of 360 video streaming using tile-based HTTP: Adaptive streaming,” in Proceedings of the 2017 ACM on Multimedia Conference. ACM, Oct. 2017, pp. 315–323.

[6] M. Xiao, C. Zhou, V. Swaminathan, Y. Liu, and S. Chen, “Bas-360: Exploring spatial and temporal adaptability in 360-degree videos over http/2,” in Proc. IEEE INFOCOM, 2018, pp. 953–961.

[7] Z. Liu, S. Ishihara, Y. Cui, Y. Ji, and Y. Tanaka, “Jet: Joint source and channel coding for error resilient virtual reality video wireless transmission,” Signal Processing, vol. 147, pp. 154–162, 2018.

[8] H. Ahmadi, O. Eltobgy, and M. Hefeeda, “Adaptive multicast streaming of virtual reality content to mobile users,” in Proceedings of the on Thematic Workshops of ACM Multimedia 2017. ACM, Oct. 2017, pp. 170–178.

[9] K. Long, C. Ye, Y. Cui, and Z. Liu, “Optimal multi-quality multicast for 360 virtual reality video,” in Proc. IEEE GLOBECOM, Dec. 2018, pp. 1–6.

[10] K. Long, C. Ye, Y. Cui, and Z. Liu, "Optimal Transmission of Multi-Quality Tiled 360 VR Video by Exploiting Multicast Opportunities." 2019 IEEE Global Communications Conference (GLOBECOM). IEEE, 2019.

[11] V. R. Gaddam, M. Riegler, R. Eg, C. Griwodz, and P. Halvorsen, “Tiling in interactive panoramic video: Approaches and evaluation,” IEEE Transactions on Multimedia, vol. 18, no. 9, pp. 1819–1831, Sept2016

[12] J. van der Hooft, S. Petrangeli, T. Wauters, R. Huysegems, P. Rondao Alface, T. Bostoen, and F. De Turck, HTTP/2-Based Adaptive Streaming of HEVC Video over 4G/LTE Networks, IEEE Communications Letters, vol. 20, no. 11, pp. 2177-2180, 2016

[13] M. Xu, C. Li, Z. Chen, Z. Wang and Z. Guan, "Assessing Visual Quality of Omnidirectional Videos," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, no. 12, pp. 3516-3530, Dec. 2019

[14] M. Viitanen, A. Koivula, A. Lemmetti, J. Vanne, and T. D. Hmlinen, “Kvazaar hevc encoder for efficient intra coding,” 2015 IEEE Inter-national Symposium on Circuits and Systems (ISCAS), May 2015, pp.1662–1665

[15] J. Li, R. Feng, Z. Liu, W. Sun and Q. Li, "Modeling QoE of Virtual Reality Video Transmission over Wireless Networks," 2018 IEEE Global Communications Conference (GLOBECOM), Abu Dhabi, United Arab Emirates, 2018

[16] A. TaghaviNasrabadi, A. Mahzari, J. D. Beshay, and R. Prakash,“Adaptive 360-degree video streaming using layered video coding,” 2017 IEEE Virtual Reality (VR), March 2017, pp. 347–348
[17] S. Rossi and L. Toni, “Navigation-aware adaptive streaming strategies for omnidirectional video,” in2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP), Oct 2017, pp. 1–6

〈発　表　資　料〉

題　　名	掲載誌・学会名等	発表年月
Joint Communication and Computational Resource Allocation for QoE-driven Point Cloud Video Streaming,	54th IEEE International Conference on Communications (ICC)	June 2020
QoE-driven Coupled Uplink and Downlink Rate Adaptation for 360-degree Video Live Streaming	IEEE Communications Letters	April 2020
Optimal Transmission of Multi-Quality Tiled 360 VR Video by Exploiting Multicast Opportunities	EEE Global Communications Conference (GLOBECOM)	December 2019