

4ZB-02

複数ユーザに対する 360 度映像伝送の低遅延化に関する一検討

岡本 翼[†] 石岡 卓将[‡] 椎名 亮太^{‡†} 福井 達也^{‡†} 小野 央也^{‡†} 藤原 稔久^{‡†}
 谷口 友宏^{‡†} 藤橋 卓也[‡] 猿渡 俊介[‡] 渡辺 尚[‡]
[†] 大阪大学工学部 [‡] 大阪大学大学院情報科学研究科
^{‡†} 日本電信電話株式会社 NTT アクセスサービスシステム研究所

1 はじめに

複数のユーザが同一の仮想現実 (Virtual Reality: VR) 空間に接続して臨場感あふれる体験を共有できるコンテンツの広がりとともに、VR 映像、すなわち 360 度映像のネットワーク伝送に対する需要が高まっている。本稿では、複数ユーザを対象とした 360 度映像伝送の低遅延化および低トラヒック化手法を提案するとともに、その効果について議論する。

2 課題

有線ネットワーク・無線ネットワークを介した 360 度映像伝送では、Dual Fisheye 形式で撮影される 360 度映像を正距円筒図法に基づいて 2 次元平面上にマッピングして 2 次元平面映像を各ユーザに伝送する。各ユーザが装着するヘッドセットは受信した 2 次元平面映像とユーザの視線位置に基づいて、2 次元平面映像のうち、一部の領域 (ビューポート) をヘッドセット上に再生する。ビューポートを介してユーザは臨場感あふれる体験に没入できる一方で、従来のネットワーク伝送方式ではユーザの VR 体験品質低下につながる 3 課題が残されている。

1 つ目の課題は Event-to-Eye (E2E) 遅延の増大にもなる体感品質の低下である。一般に、360 度映像伝送に要する E2E 遅延が 20 ms を超えたとき、VR から得られる没入感が低下すると言われている。また、E2E 遅延が 60ms を超えるとユーザは現実空間における自身の挙動と仮想空間上での自身の挙動との間にずれを感じて VR 酔いを引き起こす可能性があるとして報告されている。

E2E 遅延は大きく 360 度カメラで要する処理遅延、360 度カメラからサーバに対する 360 度映像伝送に要する遅延、サーバで要する 360 度映像の処理遅延、サーバからユーザに対する 360 度映像伝送に要する遅延、ユーザのヘッドセットで要する 360 度映像の処理遅延から構成される [1]。E2E 遅延の増大を招く 1 要因としてサーバにおいて生じる映像符号化に要する遅延の増大が挙げられる。例えば、従来の映像符号化技術である H.264/Advance Video Coding (AVC), H.265/High Efficiency Video Coding (HEVC) においてはビデオフレーム内に含まれる冗長な映像情報だけでなく、ビデオフレーム間に含まれる冗長な映像情報を削減することで 360 度映像伝送に要するトラヒックを削減する。一方で、動きベクトル推定などに挙げられる処理に起因して符号化に要する処理遅延が増大する。処理遅延の増大を抑制する 1 方法として 360 度映像を符号化せずに伝送する方法が挙げられる。一方で、後述のとおり、多大なトラヒックが必要となるため、360 度映像伝送に要する遅延が増大する。

2 つ目の課題は多大なトラヒックに起因する映像品質低下である。360 度映像は 4K, 8K, 16K に挙げられる高解像度映像である。360 度映像全体をユーザに対して伝送する場合、多大なトラヒックが必要となるため、限られた帯域下において映像品質の低下を招く。一方で、360 度映像を視聴する各ユーザは自身のヘッドセットを通して視野角に応じたビューポートを視聴する。ビューポートに含まれる領域は 360 度映像全

体の 8 分の 1 程度であるとともに、ビューポート外の映像はユーザの体感品質に影響を与えないことが分かっている。サーバはユーザごとに異なるビューポート、可用帯域を考慮して 360 度映像を配信して映像品質の向上を図る必要がある。

3 つ目の課題はユーザの視聴領域やネットワーク状況の時間変化に起因する体感品質の低下である。各ユーザのビューポートおよび可用帯域は時々刻々と変化する。ある時刻における複数ユーザのビューポート、可用帯域に応じて遠方のサーバが 360 度映像情報を符号化、伝送した場合、サーバから受信した 360 度映像情報はユーザが現在所望するビューポートと大きく異なる場合がある。このとき、ユーザはビューポートの一部を再生できず、体感品質の低下を招く可能性がある。同様に、可用帯域が時間的に変化した場合、輻輳の発生による品質低下や余剰帯域を活用できないことに起因する品質低下を招く可能性がある。本稿では、時々刻々と変化するユーザ状況のうち、ビューポートの変化、すなわち、ユーザの視線情報の変化を対象とした提案手法を設計する。

3 提案手法

本稿では、複数ユーザに対して 360 度映像を低遅延、低トラヒック、高品質に伝送するために、新たな映像伝送手法を提案する。より具体的には、上述の 3 課題を解決するために 1) タイル分割および符号化手法 JPEG-XS による低遅延符号化、2) 各ユーザの視線情報を考慮した各タイルのユニキャスト・マルチキャスト伝送、3) ユーザ環境の時間変動に対処するトランスコーディングを導入する。

3.1 全体像

図 1 に提案手法における想定環境を示す。提案手法では、360 度映像を撮影・保持するサーバ 1 台に対して複数のエッジサーバが有線を介して接続しているものとした。また、各エッジサーバには HMD を装着する複数のユーザが有線または無線を介して接続しているものとした。各ユーザは HMD から得られた視線情報を自身が接続するエッジサーバに対して逐次フィードバックする。サーバは自身に接続するエッジサーバに対してタイル分割・JPEG-XS を用いて符号化した VR 映像全体を送信する。エッジサーバはサーバから受信した VR 映像全体、ユーザから受信した視線情報に基づいて各タイルをトランスコーディングして複数ユーザに各タイルをユニキャスト・マルチキャスト伝送する。各ユーザは受信した 360 度映像全体を復号して HMD 上で再生する。

3.2 タイル分割

提案手法では、サーバが 360 度映像全体を縦方向に M 枚、横方向に N 枚のタイルに分割する。タイル分割することでユーザ環境に応じた符号化・伝送制御、並列符号化による低遅延化を可能にする。360 度映像全体の解像度を $H \times W$ 画素とすると、各タイルの大きさは $h \times w$ 画素となる。ここで h と w の取りうる値はそれぞれ $1 \leq h \leq H, 1 \leq w \leq W$ とした。

サーバおよびエッジサーバは各タイルに含まれる映像情報を JPEG-XS

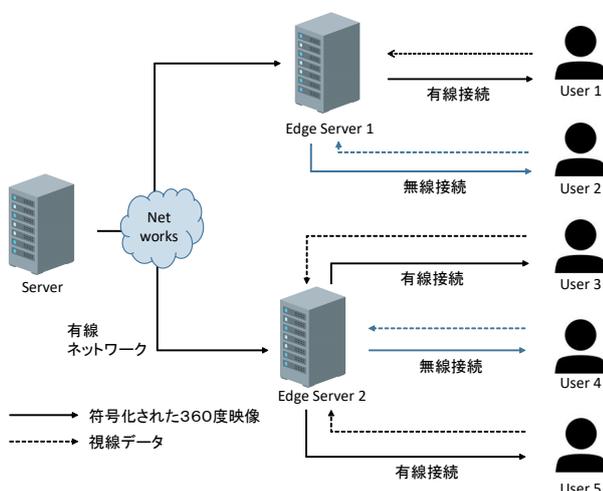


図 1: 提案手法の全体像

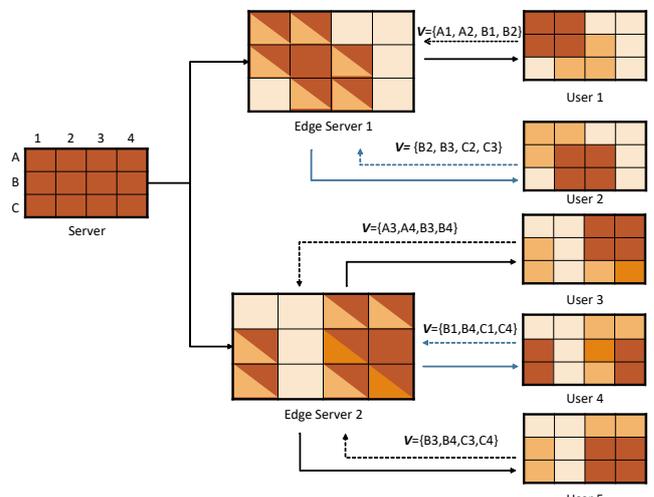


図 2: 提案手法での映像の処理

を用いて符号化する。JPEG-XS は遅延の削減に重きを置いたイントラフレームロスレス画像圧縮アルゴリズムである。JPEG-XS は従来の映像符号化技術である H.264/AVC, H.265/HEVC, 低遅延映像符号化技術 VC-2 と比較して、映像情報を低遅延に圧縮できる点、複数回の符号化（トランスコーディング）に対して耐性を持つ点から導入している。

3.3 サーバおよびエッジサーバにおけるタイル符号化および伝送手法

サーバでは、各ユーザの視線情報によらず、各タイルを共通の品質で符号化して全エッジサーバにブロードキャスト伝送する。このときの品質はサーバとエッジサーバ間のレート R にしたがって決定する。エッジサーバでは、自身に接続するユーザから受信した視線情報を元にして各ユーザの視聴領域を推定する。本稿では、ヘッドセットにおける視野角を視線位置を中心とする上下左右 90 度とみなして各ユーザの視聴領域を定め、視聴領域に含まれるすべてのタイルを視聴タイルとみなした。

提案手法では、各ユーザの視聴領域を高品質化すること、エッジサーバが推定するユーザの視聴領域とユーザが所望する視聴領域に差異があった場合、その影響を軽減することを目的としてブロードキャスト・マルチキャスト・ユニキャストを組み合わせて各タイルを伝送する。エッジサーバは各ユーザの視聴タイルを元に、全タイルを複数ユーザが共通して視聴するタイル、ユーザ 1 人が視聴するタイル、どのユーザも視聴しないタイルとに分割する。エッジサーバは複数ユーザが共通して視聴するタイルを所望する複数ユーザに対してマルチキャストすることでユーザ間で発生するタイルの冗長伝送を抑制する。一方で、エッジサーバはユーザ 1 人が所望するタイルを所望するユーザに対してユニキャスト伝送する。どのユーザも視聴しないタイルは品質を落としてエッジサーバでトランスコーディングするとともに、全ユーザに対してブロードキャストする。これはエッジサーバが受信した視線情報を元に推定した視聴領域と各ユーザが実際に所望する視聴領域との間にずれが生じた際、他のユーザが所望する視聴領域が実際の視聴領域に含まれる可能性が高いためである。

図 2 に、5 ユーザが共通のサーバに接続したときの例を示す。このうち、2 ユーザはエッジサーバ 1 に、3 ユーザはエッジサーバ 2 に有線接続または無線接続するものとした。また、360 度映像全体は $M = 3$, $N = 4$ にタイル分割するものとした。ここで各タイルの濃度が濃いほど各タイルに割り当てられたビット数が高い、すなわち、品質が高いものとした。サーバはすべてのタイルを高品質で符号化してエッジサーバ 1, エッジサーバ 2 に同時伝送する。エッジサーバ 1 に接続するユーザ 1 は全タイルのうち、A1, A2, B1, B2 のタイルを視聴領域としてエッジサーバ 1 に要求する。同様に、ユーザ 2 は B2, B3, C2, C3 のタイルを視聴領域としてエッジサーバ 1 に要求する。エッジサーバ 1 は 2 ユーザの視聴領域にしたがって共通領域となる B2 のタイルは高品質のままとしてユーザ 1・ユーザ 2 双方にマルチキャスト伝送する。また、ユーザ 1・ユーザ 2 双方が必要としないタイル A3, A4, B4, C1, C4 は低品質にトランスコーディングするとともに双方にマルチキャスト伝送する。さらに、ユーザ 1 のみが必要とするタイル A1, A2, B1 およびユーザ 2 のみが必要とするタイル B3, C2, C3 はそれぞれを低品質にトランスコーディングしたタイルを用意するとともに、低品質化したタイルを必要としないユーザ、高品質のままのタイルを必要とするユーザにユニキャスト伝送する。各ユーザは受信した各タイルをそれぞれ復号化することで高品質な視聴領域をヘッドセット上で再生できる。

4 実験評価

4.1 評価環境

提案手法の性能を評価するために、映像符号化に要する処理遅延、トラヒックに対する映像品質、視線情報の時間変化がもたらす映像品質への影響をそれぞれ評価した。符号化対象とした 360 度映像は街頭パフォーマンスをする人々の映像を利用した。360 度映像の解像度は 4096×2048 [pixel], フレームレートは 30fps とした。ビューポートの視野角は 90 度とした。360 度映像を符号化するために VC-2 と JPEG-XS を利用した。VC-2 は SMPTE 2042-1 VC-2 reference, JPEG-XS は JPEG XS Reference Software を利用した。評価環境として 2 ユーザが同時に同一の映像を視聴するものとして、そのうち 1 ユーザの映像品質を評価した。ユーザの視線情報は 30 秒間映像を視聴したユーザの視線情報のうち、10 秒間の

データを使用した。映像品質の評価指標として (1) 式に示す Viewport Normalized Peak Signal Noise-to-Ratio (VNPSNR) を用いて、ユーザの視聴領域内の映像品質を評価した。

$$VNPSNR = 10 \log_{10} \left(\frac{1}{VN MSE} \right), \quad (1)$$

$$VN MSE = \frac{1}{|V|} \sum_{(h,w) \in V} \frac{[I_y(h,w) - K_y(h,w)]^2}{I_y(h,w)^2} \quad (2)$$

ただし V は視聴領域に含まれるタイル集合、 I_y は元画像の輝度信号成分、 K_y は復元画像の輝度信号成分を表す。

4.2 タイル分割が符号化遅延にもたらす影響

本節では、タイル分割数に起因して各タイルの符号化・復号化に要する遅延が受ける影響を評価した。図 3 にタイル分割数に対して 1 タイルが要する符号化遅延および復号化遅延を示す。ここで、全タイルは並列に符号化するものとした。評価結果からタイル数が増加する、すなわち、タイルあたりに含まれる画素数が小さくなるにつれて符号化に要する遅延および復号化に要する遅延がそれぞれ小さくなるのが分かった。また、360 度映像全体を 1000 タイルに分割して並列実行した場合においても、その遅延量は著しく増加しないことが分かった。

4.3 トラヒックに対する映像品質

本節では、360 度映像のうち、1 ビデオフレームを伝送するのにかかるトラヒックに対する視聴領域内の映像品質を評価した。図 4 に 1 ビデオフレームあたりのトラヒックに対する VNPSNR を示す。比較手法として、JPEG-XS を用いてビデオフレーム全体を同一品質で伝送する uniform (JPEG-XS), VC-2 を用いて映像全体を同一品質で符号化して伝送する uniform (VC-2), VC-2 を用いて視聴領域を高品質で伝送する Adaptive (VC-2) を用いた。トラヒックとしてエッジサーバとユーザ間に発生するトラヒックを評価した。ここでエッジサーバとユーザ間には 1 フレーム分の遅延が発生するものとした。

評価結果から提案手法がトラヒックに関わらず、視聴領域に対して最も優れた品質を達成できていることが分かった。また、JPEG-XS を導入することで低遅延化を達成しつつ、より高い品質を達成できることが分かった。

4.4 視線情報の時間変化に起因する映像品質への影響

本節では、サーバとエッジサーバ間の遅延量が増加したとき、すなわち、サーバが各ユーザの過去の視線情報を元に各タイルの品質を適応的に符号化した場合の映像品質を評価した。図 5 に、サーバとエッジサーバ間の遅延量に対する VNPSNR を示す。比較手法として、視線情報に基づいてサーバが JPEG-XS を用いて各タイルを符号化する server-based, VC-2 を用いてエッジサーバが各タイルをトランスコーディングする edge-based (VC-2) を導入した。ここで、server-based では過去の視線情報をそのまま使用するものとした。

評価結果から server-based は遅延が大きくなるにつれて映像品質が低下してしまうこと、提案手法はサーバエッジサーバ間の遅延の影響を受けず、ユーザの視線情報に基づいて高品質化した視聴領域を伝送できることが分かった。

5 おわりに

本研究では複数ユーザを対象とした 360 度映像伝送の低遅延化および低トラヒック化手法を提案した。評価結果から、同程度のトラヒック下において提案手法はより高い品質を達成できることが分かった。

謝辞

本研究は JSPS 科研費 (JP19H01101, 20K19783), NTT アクセスサービスシステム研究所の支援の下で行った。

参考文献

[1] J. Yi, M. R. Islam, S. Aggarwal, D. Koutsonikolas, Y. Charlie Hu, and Z. Yan, "An Analysis of Delay in Live 360° Video Streaming Systems," in Proceedings of the 28th ACM International Conference on Multimedia, pp. 982–990, 2020.

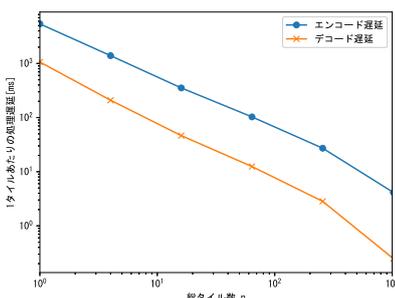


図 3: タイル分割数に対する処理遅延

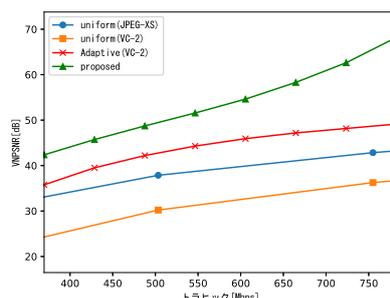


図 4: トラヒックに対する映像品質

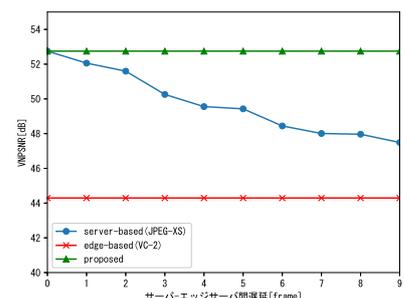


図 5: 遅延に対する映像品質