Optimal Streaming of 360 VR Videos with Perfect, Imperfect and Unknown FoV Viewing Probabilities

Presenter: Lingzhi Zhao
Author: Lingzhi Zhao*, Ying Cui*, Chengjun Guo* and Zhi Liu†

Shanghai Jiao Tong University*   Shizuoka University†

IEEE GLOBECOM 2020

Outline

Outline

Background

- A 360 VR video is generated by capturing a scene of interest in every direction at the same time using omnidirectional cameras
- A user wearing a VR headset or head mounted display (HMD) can freely watch the scene of interest in any viewing direction at any time
- VR has vast applications in entertainment, education, medicine, etc.
- Data size of 360 VR video is up to 16 times that of traditional video
- At any moment, a VR user is interested in only one viewpoint, i.e., the center of one part of the 360 VR video, referred to as *FoV*

Background

- To improve transmission efficiency, a 360 VR video is divided into smaller rectangular segments of the same size, referred to as *tiles*, and the set of tiles covering the predicted FoV or FoVs with higher viewing probabilities are transmitted
- Pre-encode each tile into multiple representation with different quality levels allows quality adaptation according to user heterogeneity
  - e.g., different cellular usage costs, display resolutions of devices, channel conditions, etc.

Previous work

- Consider wireless streaming of one [Long'18][Long'20][Zou'20] or multiple [Huang'18][Yang'19] multi-quality tiled 360 VR videos, and optimize the quality level selection and communication resource allocation
  - maximize the total utility [Long'18] [Huang'18][Yang'19]
  - minimize the total distortion [Zou'20]
  - minimize the total transmission power [Long'20]
- Assume that FoVs that may be watched are perfectly known [Long'18][Long'20] [Zou'20][Huang'18][Yang'19], or that the viewing probability distributions over those FoVs are perfectly known [Zou'20] [Huang'18][Yang'19]
  - lead to poor performance when prediction errors are large
- Consider a single-antenna server
  - cannot exploit spatial degrees of freedom to provide performance
- Our goal is to fix the above two limitations

Outline

Multi-quality tiled 360 VR video model

- Divide a 360 VR video into $X \times Y$ tiles
- Consider $\bar{I}$ viewpoints ($\bar{I}$ FoVs), and denote $\bar{\mathcal{I}} \triangleq \{1, \dots, \bar{I}\}$
- $\mathcal{F}_i$ denotes the set of tiles included in the i-th FoV, where $i \in \bar{\mathcal{I}}$
- Pre-encode each tile into L representations corresponding to L quality levels, and denote $\mathcal{L} \triangleq \{1, \dots, L\}$
- D_l denotes encoding rate (in bits/s) of l-th representation of a tile, where $l \in \mathcal{L}$, $D_1 < D_2 ... < D_L$
- Consider wireless streaming of K multi-quality tiled 360 VR videos from a server to K users, respectively, and denote $\mathcal{K} \triangleq \{1, \dots, K\}$
- $\mathcal{I}_k$ represents the sets of indices of the I_k FoVs of video k that may be watched by user k (known to the server), where $k \in \mathcal{K}$
- $p_{i,k}$ denotes the probability that i-th FoV of video k is viewed by user k, where $i \in \mathcal{I}_k$, $k \in \mathcal{K}$, and let $\mathbf{p}_k \triangleq (p_{i,k})_{i \in \mathcal{I}_k}$, $p_{i,k} \geq 0$ and $\sum_{i \in \mathcal{I}_k} p_{i,k} = 1$

Three cases of FoV viewing probability distributions

- Perfect FoV viewing probability distributions:
  - the exact values of $\mathbf{p}_k, k \in \mathcal{K}$ are assumed to be known to the server
- Imperfect FoV viewing probability distributions:
  - the server knows that $\mathbf{p}_k \in \mathcal{P}_k, k \in \mathcal{K}$, where

$$\mathcal{P}_k \triangleq \left\{ \mathbf{p}_k \;\middle|\; \underline{p}_{i,k} \leq p_{i,k} \leq \overline{p}_{i,k}, i \in \mathcal{I}_k, \sum_{i \in \mathcal{I}_k} p_{i,k} = 1 \right\}, k \in \mathcal{K}.$$

with $\underline{p}_{i,k} \triangleq \max\{\hat{p}_{i,k} - \varepsilon_{i,k}, 0\}$ and $\overline{p}_{i,k} \triangleq \min\{\hat{p}_{i,k} + \varepsilon_{i,k}, 1\}$

- Unknown FoV viewing probability distributions:
  - there is no information about $\mathbf{p}_k, k \in \mathcal{K}$

Encoding rates of tiles and FoVs

- Tiles in $\overline{\mathcal{F}}_k \triangleq \cup_{i\in\mathcal{I}_k} \mathcal{F}_i$ may be transmitted to user $k \in \mathcal{K}$
- $R_{x,y,k}$ denotes the encoding rate of the (x,y)-th tile of video k

$$R_{x,y,k} \in \{0, D_1, \dots, D_L\},\ (x,y) \in \overline{\mathcal{F}}_k,\ k \in \mathcal{K}$$

- $r_{i,k}$ represents the minimum of the encoding rates of the tiles in the i-th FoV of video k (minimum encoding rate of the i-th FoV of video k)

$$r_{i,k} \in \{0, D_1, \dots, D_L\},\ i \in \mathcal{I}_k,\ k \in \mathcal{K}$$

- Consider relative smoothness requirement

$$r_{i,k} \leq R_{x,y,k} \leq r_{i,k} + \delta,\ (x,y) \in \mathcal{F}_i,\ i \in \mathcal{I}_k,\ k \in \mathcal{K}$$

Performance metrics

- Let $U(r)$ denote the utility for an FoV with the minimum encoding rate r
  - $U(\cdot)$ can be any nonnegative, strictly increasing and strictly concave function, and $U(0) = 0$
- In the case of perfect FoV viewing probability distributions (i.e., case-pp), the performance metric is the *average total utility*:
  - $Q^{(\text{pp})}(\mathbf{r}) = \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{I}_k} p_{i,k} U(r_{i,k})$
- In the case of imperfect FoV viewing probability distributions (i.e., case-ip), the performance metric is the *worst average total utility*:
  - $Q^{(\text{ip})}(\mathbf{r}) = \sum_{k \in \mathcal{K}} \min_{\mathbf{p}_k \in \mathcal{P}_k} \sum_{i \in \mathcal{I}_k} p_{i,k} U(r_{i,k})$
- In the case of unknown FoV viewing probability distributions (i.e., case-up), the performance metric is the *worst total utility*:
  - $Q^{(\text{up})}(\mathbf{r}) = \sum_{k \in \mathcal{K}} \min_{i \in \mathcal{I}_k} U(r_{i,k})$

Illustration example

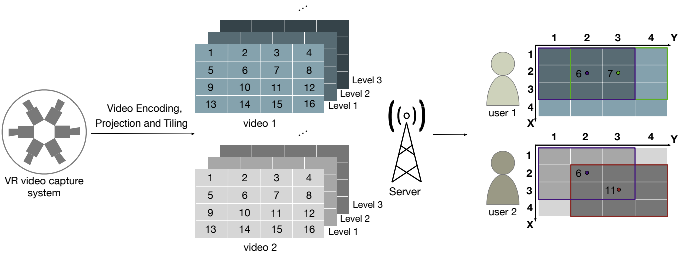

Figure: Illustration example

- $X = 4,\ Y = 4,\ \bar{I} = 16,\ L = 3,\ K = 2,\ \mathcal{I}_1 = \{6, 7\},\ \mathcal{I}_2 = \{6, 11\}$
- $\mathcal{F}_6 = \{(1,1), (1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2), (3,3)\}$
- $\mathcal{F}_7 = \{(1,2), (1,3), (1,4), (2,2), (2,3), (2,4), (3,2), (3,3), (3,4)\}$
- $r_{6,1} = D_2,\ r_{7,1} = D_1,\ R_{x,y,1} = D_2, (x, y) \in \mathcal{F}_6,\ R_{x,y,1} = D_1, (x, y) \in \mathcal{F}_7 \backslash \mathcal{F}_6$
- $\mathcal{F}_6 = \{(1,1), (1,2), (1,3), (2,1), (2,2), (2,3), (3,1), (3,2), (3,3)\}$
- $\mathcal{F}_{11} = \{(2,2), (2,3), (2,4), (3,2), (3,3), (3,4), (4,2), (4,3), (4,4)\}$
- $r_{6,2} = D_2, r_{11,2} = D_3, R_{x,y,2} = D_3, (x, y) \in \mathcal{F}_{11}, R_{x,y,2} = D_2, (x, y) \in \mathcal{F}_6 \backslash \mathcal{F}_{11}$

Physical layer model

- Consider one M-antennas server and K single-antenna users
- Consider N subcarriers, each with the bandwidth B
- Assume block fading
  - i.e., the channel on each subcarrier remains constant over the considered time duration
- $\mathbf{h}_{k,n}^H \in \mathbb{C}^{1\times M}$ denotes the downlink channel vector on subcarrier n between user k and the server
- Assume that the channel state information is perfectly known at the server and the users

Rate splitting scheme

- Encoded (source coding) bits of the tiles in $\overline{\mathcal{F}}_k$ that will be transmitted to user $k \in \mathcal{K}$ are "aggregated" into one message
- The aggregated message for user $k \in \mathcal{K}$ is split into a common part of rate $d_{c,k}$ and a private part of rate $d_{p,k}$

$$\sum_{(x,y)\in\overline{\mathcal{F}}_k} R_{x,y,k} = d_{c,k} + d_{p,k},\ \ k \in \mathcal{K}$$

  - the common parts of the messages of the K users are combined into a *common message*
  - the private part of user k's message is referred to as user k's *private message*
- The common message and the K users' private messages are encoded (channel coding) into codewords that span over N subcarriers
- $\mathbf{w}_{c,n}$ and $\mathbf{w}_{k,n}$ denote the common beamforming vector and the private beamforming vector for user k on subcarrier n.
- Total transmission power constraint

$$\sum_{n\in\mathcal{N}} \left(||\mathbf{w}_{c,n}||_2^2 + \sum_{k\in\mathcal{K}} ||\mathbf{w}_{k,n}||_2^2 \right) \leq P$$

Successive decoding

- User $k \in \mathcal{K}$ decodes the common message by treating the interference from the K users' private messages on each subcarrier as noise, and removes it
- User $k \in \mathcal{K}$ decodes his private message by treating the interference from the remaining $K-1$ users' private messages on each subcarrier as noise
- Successful transmission constraints

$$\sum_{k \in \mathcal{K}} d_{c,k} \leq \sum_{n \in \mathcal{N}} B \log_2 \left(1 + \frac{|\mathbf{h}_{k,n}^H \mathbf{w}_{c,n}|^2}{\sum_{j \in \mathcal{K}} |\mathbf{h}_{k,n}^H \mathbf{w}_{j,n}|^2 + \sigma^2} \right)$$

$$d_{p,k} \leq \sum_{n \in \mathcal{N}} B \log_2 \left(1 + \frac{|\mathbf{h}_{k,n}^H \mathbf{w}_{k,n}|^2}{\sum_{j \in \mathcal{K}, j \neq k} |\mathbf{h}_{k,n}^H \mathbf{w}_{j,n}|^2 + \sigma^2} \right)$$

Outline

Problem formulation

Problem 1 (Total Utility Maximization) $t = \text{pp,ip,up}$

$$U^{(t)\star} \triangleq \max_{\mathbf{R},\mathbf{r},\mathbf{d},\mathbf{w}} \quad Q^{(t)}(\mathbf{r})$$

$$\begin{aligned}
\text{s.t.}\quad & 0 \le R_{x,y,k} \le D_L,\ (x,y) \in \overline{\mathcal{F}}_k,\ k \in \mathcal{K} \ \text{(continuous relaxation)} \\
& 0 \le r_{i,k} \le D_L,\ i \in \mathcal{I}_k,\ k \in \mathcal{K} \ \text{(continuous relaxation)} \\
& r_{i,k} \le R_{x,y,k} \le r_{i,k} + \delta,\ (x,y) \in \mathcal{F}_i,\ i \in \mathcal{I}_k,\ k \in \mathcal{K} \\
& \sum_{(x,y)\in\overline{\mathcal{F}}_k} R_{x,y,k} = d_{c,k} + d_{p,k},\ k \in \mathcal{K} \\
& \sum_{n\in\mathcal{N}} \left(||\mathbf{w}_{c,n}||_2^2 + \sum_{k\in\mathcal{K}} ||\mathbf{w}_{k,n}||_2^2 \right) \le P \\
& \sum_{k\in\mathcal{K}} d_{c,k} \le \sum_{n\in\mathcal{N}} B\log_2\left(1 + \frac{|\mathbf{h}_{k,n}^H \mathbf{w}_{c,n}|^2}{\sum_{j\in\mathcal{K}} |\mathbf{h}_{k,n}^H \mathbf{w}_{j,n}|^2 + \sigma^2}\right) \\
& d_{p,k} \le \sum_{n\in\mathcal{N}} B\log_2\left(1 + \frac{|\mathbf{h}_{k,n}^H \mathbf{w}_{k,n}|^2}{\sum_{j\in\mathcal{K},j\neq k} |\mathbf{h}_{k,n}^H \mathbf{w}_{j,n}|^2 + \sigma^2}\right), k \in \mathcal{K}
\end{aligned}$$

- It is a challenging non-convex problem with a non-differentiable objective function ($t = \text{ip,up}$)

Optimality properties

For all $i \in \mathcal{I}_k, k \in \mathcal{K}$, define $\mathcal{T}_{i,k} \triangleq \mathcal{F}_i \backslash \overline{\mathcal{F}}_k$.

Theorem 1 (Optimality Properties)

(i) For $t =$ pp, ip, up, $R^{(t)\star}_{x,y,k} = \max_{i \in \mathcal{I}_k:(x,y)\in\mathcal{F}_i} r^{(t)\star}_{i,k}, (x,y) \in \overline{\mathcal{F}}_k,\ k \in \mathcal{K}$.

(ii) For $t =$ pp and for all $i,j \in \mathcal{I}_k, i \neq j, k \in \mathcal{K}$, if $p_{i,k} \leq p_{j,k}$, and $|\mathcal{T}_{i,k}| > |\mathcal{T}_{j,k}| > 1$, then $r^{(\text{pp})\star}_{i,k} \leq r^{(\text{pp})\star}_{j,k}$. For $t =$ ip and for all $i,j \in \mathcal{I}_k, i \neq j, k \in \mathcal{K}$, if $\overline{p}_{i,k} \leq \underline{p}_{j,k}$, and $|\mathcal{T}_{i,k}| > |\mathcal{T}_{j,k}| > 1$, then $r^{(\text{ip})\star}_{i,k} \leq r^{(\text{ip})\star}_{j,k}$. For $t =$ up, $r^{(\text{up})\star}_{i,k}, i \in \mathcal{I}_k$ are identical, $k \in \mathcal{K}$.

(iii) $U^{(\text{pp})\star} \geq U^{(\text{ip})\star} \geq U^{(\text{up})\star}$.

- (i) Indicates that at least one of $r_{i,k} \leq R_{x,y,k}$ is active at an optimal solution
- (ii) In case-pp and case-ip, an FoV with a higher viewing probability has a higher minimum encoding rate; in case-up, the minimum encoding rates are identical
- (iii) Shows relationship among the optimal values of Problem 1 for the three cases

Outline

Solution

- Transform Problem 1 into an equivalent DC programming problem with a differentiable objective function in each case
  - in case-pp, introduce auxiliary variables and extra constraints
  - in case-ip, replace the inner problem with its dual problem, and introduce auxiliary variables and extra constraints
  - in case-up, cast in hypograph form, and introduce auxiliary variables and extra constraints
- Obtain a stationary point by CCCP [Sun'17] in each case
  - main idea: solve a sequence of successively refined approximate convex problems

Outline

Simulation setup

- Consider five 360 VR video sequences, i.e., *Diving*, *Rollercoaster*, *Timelapse*, *Venice*, *Paris*, provided by [Corbillon'17]
- Set $X = 8$, $Y = 8$, $\bar{I} = 64$, $K = 5$

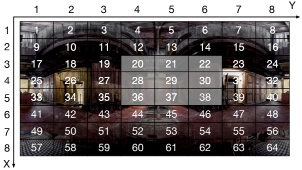

Figure: Illustration of tiles, viewpoints and FoVs in a 360 VR video. The grey area represents FoV 29 which is centered at viewpoint 29.

Table: Encoding rates (in Mbit/s) for $L = 3, 5, 7$.

L	$D_l,\ l \in \mathcal{L}$.
3	$D_1 = 14.46,\ D_2 = 52.97,\ D_3 = 87.75.$
5	$D_1 = 14.46,\ D_2 = 37.10,\ D_3 = 52.97,\ D_4 = 69.53,\ D_5 = 87.75.$
7	$D_1 = 14.46,\ D_2 = 37.10,\ D_3 = 46.20,\ D_4 = 52.97,\ D_5 = 59.45,\ D_6 = 69.53,\ D_7 = 87.75.$

Simulation setup

- For each video sequence, obtain a viewpoint sequence for each user, with one viewpoint for every 1/3 second, based on the viewpoint data of 59 users provided by [Corbillon'17]
- View users 2, 8, 24, 32 and 40 as the users who request videos 1, 2, 3, 4 and 5, respectively
- Consider the prediction of the 3-rd element in user k's viewpoint sequence given the 2-nd element
- i_k denotes the index of the FoV of user k corresponding to the 2-nd element
- Set $\mathcal{I}_k = \{i_k - 8, i_k - 1, i_k, i_k + 1, i_k + 8\}$
- $n_{i_k,i}$ denotes number of users with the 2-nd element and 3-rd element in his viewpoint sequence for video k being i_k and i, respectively
- Calculate the FoV viewing probabilities according to $p_{i,k} = \frac{n_{i_k,i}}{\sum_{i \in \mathcal{I}_k} n_{i_k,i}}, i \in \mathcal{I}_k, k \in \mathcal{K}$

Simulation setup

Table: Prediction parameters.

k	Video sequence	User	Current FoV	Predicted FoVs	FoV viewing probability distributions
1	*Diving*	2	28	$\mathcal{I}_1 = \{20, 27, 28, 29, 36\}$	$(p_{20,1}, p_{27,1}, p_{28,1}, p_{29,1}, p_{36,1}) = (0.4138, 0.1724, 0.2414, 0.1667, 0.0417)$
2	*Rollercoaster*	8	21	$\mathcal{I}_2 = \{13, 20, 21, 22, 29\}$	$(p_{13,2}, p_{20,2}, p_{21,2}, p_{22,2}, p_{29,2}) = (0, 0.4615, 0.3077, 0.0769, 0.1538)$
3	*Timelapse*	24	24	$\mathcal{I}_3 = \{16, 23, 24, 17, 32\}$	$(p_{16,3}, p_{23,3}, p_{24,3}, p_{17,3}, p_{32,3}) = (0.1481, 0.037, 0.2963, 0.5185, 0)$
4	*Venice*	32	29	$\mathcal{I}_4 = \{21, 28, 29, 30, 37\}$	$(p_{21,4}, p_{28,4}, p_{29,4}, p_{30,4}, p_{37,4}) = (0.25, 0.375, 0.25, 0.0625, 0.0625)$
5	*Paris*	40	18	$\mathcal{I}_5 = \{10, 17, 18, 19, 26\}$	$(p_{10,5}, p_{17,5}, p_{18,5}, p_{19,5}, p_{26,5}) = (0.375, 0.5, 0.125, 0, 0)$

- Set $\hat{p}_{i,k} = p_{i,k}, \varepsilon_{i,k} = \varepsilon, i \in \mathcal{I}_k, k \in \mathcal{K}$
- Set $B = 1\text{MHz}$, $N = 128$, $P = 30$ dBm and $\sigma^2 = 10^{-9}$ W
- Consider $U(r) = 0.6 \log(1000 \frac{r}{D_L})$ [Zhang'13]
- Choose $\mathbf{h}_{k,n}$ according to $\mathcal{CN}(0, \mathbf{1}_{M\times M})$, evaluate the average performance metrics over 100 random realizations of $\mathbf{h}_{k,n}$

Numerical results of proposed solutions

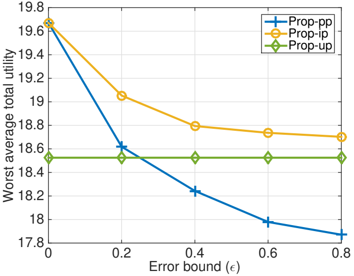

Figure: Worst average total utility comparision of the proposed solutions at $M = 64$.

- The worst average total utility of Prop-up is irrelevant to ε
- In case-ip, the worst average total utility of Prop-ip is greater than those of Prop-pp and Prop-up
  - reveal the importance of explicitly considering imperfectness of the predicted FoV viewing probability distributions in this case

Numerical results of proposed solutions

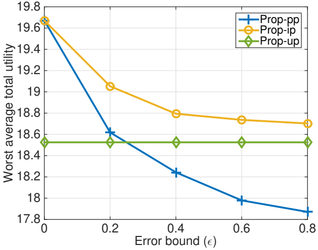

Figure: Worst average total utility comparision of the proposed solutions at $M = 64$.

- In case-ip, the worst-case average total utility of Prop-up is greater than that of Prop-pp when ε is large
    - Prop-up is designed to maximize the worst total utility and does not depend on information of FoV viewing probability distributions
- The gain of Prop-ip over Prop-pp increases with ε
    - it is more important to consider FoV prediction error when ε is large
- The gain of Prop-ip over Prop-up decreases with ε
    - imperfect FoV viewing probability distributions become less important when ε is large

Numerical results of proposed solutions

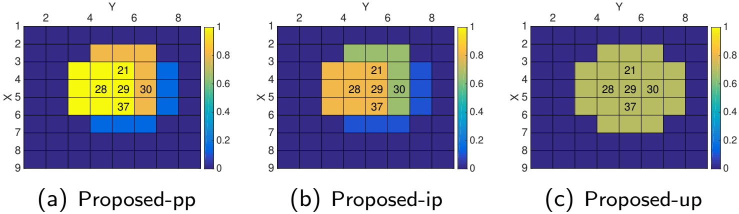

Figure: Encoding rates of the tiles for user 4 given by Prop-pp, Prop-ip and Prop-up at $M = 64$, $\varepsilon = 0.4$.

- The encoding rates of the tiles in an FoV with a larger viewing probability are higher
- The encoding rates of the tiles given by Prop-up are identical
- Such observations are in accordance with the optimality properties in Statement (ii) of Theorem 1

Baseline schemes

- BL-SDMA-OPT
  - SDMA is adopted, and beamforming optimization is considered on each subcarrier
  - a suboptimal solution of Problem 1 with $d_{c,k} = 0, k \in \mathcal{K}$ and $\mathbf{w}_{c,n} = 0, n \in \mathcal{N}$ (which is a DC programming) is obtained using CCCP
- BL-SDMA-ZF
  - SDMA is adopted, and zero forcing beamforming is adopted on each subcarrier
  - an optimal solution of Problem 1 with the zero forcing beamformers (which is convex) is obtained by standard convex optimization
- BL-OFDMA-MRT
  - OFDMA is adopted, and the maximum ratio transmission (MRT) is adopted for each user
  - an optimal solution of Problem 1 with the MRT beamformers is obtained by continuous relaxation and the KKT conditions

Numerical results of proposed solutions and baseline schemes

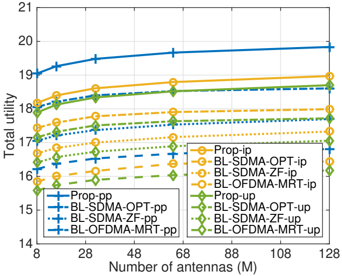

Figure: Total utility comparision at $\varepsilon = 0.4$.

In each case:

- The total utility of each scheme increases with M

Numerical results of proposed solutions and baseline schemes

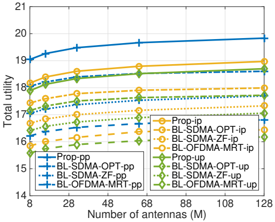

Figure: Total utility comparision at $\varepsilon = 0.4$.

In each case:

- The proposed solution outperforms BL-SDMA-OPT
  - the cost to suppress interference in BL-SDMA-OPT can be high
  - rate splitting scheme partially decodes interference and partially treats interference as noise
- The proposed solution outperforms BL-SDMA-ZF
  - the cost to suppress interference in BL-SDMA-ZF can be high
  - beamforming directions are optimized in rate splitting scheme
- The proposed solution outperforms BL-OFDMA-MRT
  - rate splitting scheme has effective spatial multiplexing

Outline

Conclusion

- Consider optimal wireless streaming of multi-quality tiled 360 VR videos from a multi-antenna server to multiple single-antenna users in a multi-carrier system
- To capture the impact of FoV prediction, consider perfect, imperfect and unknown FoV viewing probability distributions, and user the average total utility, worst average total utility and worst total utility as the respective performance metrics
- To improve transmission efficiency for multiple users, adopt rate splitting with successive decoding
- In each case, opimize the rate adaptation, resource allocation and beamforming to maximize the total utility
- In each case, obtain a stationary point using CCCP
- Numerical results demonstrate the proposed solutions achieve notable gains over existing schemes in all three cases and reveal the impact of FoV prediction and its accuracy on wireless streaming of multi-quality tiled 360 VR videos

Related publications

- Journal papers
  - "Optimal wireless streaming of multi-quality 360 VR videos with perfect, imperfect and unknown FoV viewing probabilities," submitted to *IEEE TIP*, 2020.
  - "Optimal wireless streaming of multi-quality 360 VR video by exploiting natural, relative smoothness-enabled and transcoding-enabled multicast opportunities," *IEEE Trans. Multimedia*, 2020.
  - "Optimal multicast of tiled 360 VR video," *IEEE Wireless Commun. Lett.*, 2019
  - "Optimal multicast of tiled 360 VR video in OFDMA systems," *IEEE Commun. Lett.*, 2018
- Conference papers
  - "Optimal transmission of multi-quality tiled 360 VR video by exploiting multicast opportunities," in *Proc. of IEEE GLOBECOM*, 2019
  - "Optimal multi-quality multicast for 360 virtual reality video," in *Proc. of IEEE GLOBECOM*, 2018

Reference I

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

[2] K. Long, Y. Cui, C. Ye, and Z. Liu, "Optimal wireless streaming of multi-quality 360 VR video by exploiting natural, relative smoothness-enabled and transcoding-enabled multicast opportunities," *IEEE Trans. Multimedia*, 2020.

[3] J. Zou, C. Li, C. Liu, Q. Yang, H. Xiong, and E. Steinbach, "Probabilistic tile visibility-based server-side rate adaptation for adaptive 360-degree video streaming," *IEEE JSTSP*, vol. 14, no. 1, pp. 161–176, Jan. 2020.

[4] W. Huang, L. Ding, G. Zhai, X. Min, J.-N. Hwang, Y. Xu, and W. Zhang, "Utility-oriented resource allocation for 360-degree video transmission over heterogeneous networks," *Digit. Signal Process.*, vol. 84, pp. 1–14, Oct. 2018.

[5] J. Yang, J. Luo, D. Meng, and J. Hwang, "QoE-driven resource allocation optimized for delay-sensitive VR video uploading over cellular network," in *Proc. of IEEE ISCC*, Jun. 2019, pp. 1–6.

[6] A. Wiesel, Y. C. Eldar, and S. Shamai, "Zero-forcing precoding and generalized inverses," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4409–4418, Sep. 2008.

[7] H. Joudeh and B. Clerckx, "Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: a rate-splitting approach," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4847–4861, Nov. 2016.

Reference II

[8] J. Chakareski, R. Aksu, X. Corbillon, G. Simon, and V. Swaminathan, "Viewport-driven rate-distortion optimized 360° video streaming," in *Proc. of IEEE ICC*, May. 2018, pp. 1–7.

[9] L. Zhao, Y. Cui, C. Guo, Z. Liu and S. Yang "Optimal wireless streaming of multi-quality 360 VR videos with perfect, imperfect and unknown FoV viewing probabilities," *Technical Report*, 2020. [Online]. Available: http://iwct.sjtu.edu.cn/personal/yingcui/publicans.html

[10] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2017.

[11] X. Corbillon, F. De Simone, and G. Simon, "360-degree video head movement dataset," in *Proc. of ACM MMSys*, June 2017, pp. 199–204.

[12] W. Zhang, Y. Wen, Z. Chen, and A. Khisti, "QoE-driven cache management for HTTP adaptive bit rate streaming over wireless networks," *IEEE Trans. Multimedia*, vol. 15, no. 6, pp. 1431–1445, Oct. 2013.

Thank you for listening!