ICC 2021 - IEEE International Conference on Communications | 978-1-7281-7122-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICC42927.2021.9500862

Cooperative 360° Video Delivery Network: A Multi-Agent Reinforcement Learning Approach

Fenghe Hu, Yansha Deng, and A. Hamid Aghvami
Department of Engineering, King's College London, UK
E-mail:fenghe.hu, yansha.deng, hamid.aghvami@kcl.ac.uk.

Abstract—**With the stringent requirement of receiving video from unmanned aerial vehicle (UAV) from anywhere in the stadium of sports events and the significant-high per-cell throughput for video transmission to virtual reality (VR) users, a promising solution is a cell-free multi-group broadcast (CF-MB) network with cooperative reception and broadcast access points (AP). To explore the benefit of broadcasting user-correlated decode-dependent video resources to spatially correlated VR users, the network should dynamically cluster APs into virtual cells for a different group of VR users with overlapped video requests. We first introduce the conventional non-learning-based association algorithms. We then formulate the association problem into a networked-distributed Partially Observable Markov decision process (ND-POMDP). To solve it, we propose a multi-agent deep DRL algorithm based on the rainbow agent with a convolutional neural network (CNN) to generate decisions from observation. Our simulation results shown that our CF-MB network can effectively handle real-time video transmission from UAVs to VR users. Our proposed learning architectures is effective and scalable for a high-dimensional cooperative association problem with increasing APs and VR users. Also, our proposed algorithms outperform non-learning based methods with significant performance improvement.**

I. Introduction

Unmanned aerial vehicle (UAV) systems bring fast and easy accessibility of aerial video capture into our daily life. A typical application is the True View Technology for large sports events introduced by Intel, which enhances the audiences' viewing experience by allowing audiences to customize their viewing angles freely and be immersed in a selected environment with the help of a large camera array distributed around the stadium and head-mounted displays (HMD) for virtual reality (VR) audiences. To realise the full vision of event enhancing VR video capture with highly mobile UAV, a wireless network is needed to receive, process and transmit the captured 360° VR video from multiple UAVs to massive VR users. However, the overall capacity requirement for such service from network to massive VR users can reach $22Tbps/km^2$ level [1]. Also, this service requires seamless real-time responses to VR users' viewpoint selections, where the newly generated video frames should be successfully transmitted and decoded without noticeable jitter or delay [1].

Existing research on VR video transmission has been mainly focused on reducing the transmission delay via caching and wireless resource allocation [2]. In [2], the authors optimized the resource allocation for VR video transmission under the consideration of data correlation. However, [2], [3] assumed pre-stored independent VR video resource in the form of chunk or image without considering task correlation and video increment decoding schemes.

To tackle the real-time VR video transmission from UAVs to a large number of VR users with content request correlation, the broadcasting of the same content to VR users with the same request is shown to be a promising solution [3], [4], which largely reduce the bandwidth requirement. However, the limited coverage and inter-cell interference are detrimental to the broadcasting system, especially for cell-edge VR users. One possible solution is the cooperative transmission, which has been proposed in [5]. By introducing the concept of cooperative transmission into a large scale network. The authors in [5], [6] proposed a user-centric cell-free (CF) multi-input-multi-output (MIMO) network to facilitate wide range cooperation among a large number of distributed access points (AP) for multiple user groups with cooperative transmission by a central server via high-speed backhaul links. However, a large number of VR users and cooperative APs introduce a association problem with high-dimensional environment (i.e. channel state, actions of multiple AP). Luckily, a convolutional neural network (CNN)-based DRL and multi-agent setting are known to be able to extract complex wireless features from the environment [7].

Motivated by the above, we first propose a decode-forward (DF) CF-MB network for VR video resource transmission with *UAV-APs* uplink transmission from

978-1-7281-7122-7/21/$31.00 ©2021 IEEE

UAV camera to APs group, and *APs-VR* downlink transmission from APs group to users. We also define our VR video resource via tiles, and QoE metric via the viewpoint-peak-signal-noise-ratio (V-PSNR) based on the number of successfully decoded tiles at the VR users' sides. Then, we formulate our optimization problem as the maximization of the total V-PSNR in each group of picture (GOP) for all the VR users. To dynamically associate APs, we formulate the association problem as a networked-distributed Partially Observable Markov Decision Process (ND-POMDP), which are coordinated via mean-field theorem. We then propose a federated distributed multi-agent DRL approach with the distributed rainbow agent at each AP. Our results shown that our distributed algorithm outperform the non-learning-based baselines, and can efficiently solve the association optimization problem while adapting to the dynamic environment with arbitrary density, locations, and request patterns VR users.

The remainder of this paper is organized as follows. Section II illustrates the communication model and video decoding model. And we define our optimization target with defining viewpoint peak signal-noise ratio (V-PSNR) as the QoE metric. In Section III, we introduce the ND-POMDP problem, which is then solved a federated multi-agent association setting. Then, in Section IV, the numerical results are presented. Finally, we conclude the paper in Section V.

II. System Model

As illustrated in Fig. 1, we consider a CF-MB network for 360° video transmission in a large sports event. This CF-MB network is composed of 1) a set of APs $\mathcal{B}$, which are located in the grid; 2) a central server, which connects all APs through backhaul optical links; 3) a set of randomly located camera UAVs $\mathcal{U}$, where each UAV provides the video resource from their orientation; and 4) a set of VR $\mathcal{V}$ users, whose locations follow Poisson cluster process (PCP) with $|\mathcal{U}|$ clusters. Considering that the VR users' video requests can correlate to their location in large sports event scenarios, we apply PCP distribution to capture the geographically correlated video requests. We assume that VR users in each cluster request video resources from the same UAV, while the clusters can be overlapped or disjointed. All nodes remain spatially static for each group-of-picture (GOP) once deployed. The network acts as a DF network, which receives the video from UAV (*UAV-APs* uplink) and broadcasts to target the VR user group (*APs-VR* downlink).

A. Transmission Channel Model

To capture the channel characteristic for UAVs and VR user groups, we consider different channel models for the *UAV-APs* uplink and the *APs-VR* downlink, respectively. We assume that both channels follow block fading assumption. The *UAV-APs* uplink from UAV to APs and APs-UAV downlink from APs to VR user group occupy B_{UL} and B_{DL} bandwidth, respectively. We also assume a perfect channel state is available at the APs.

1) UAV-APs Uplink

The *UAV-APs* uplink between a UAV and a AP group forms a virtual single-input-multi-output (SIMO) system, where multiple APs are associated to enhance the signal reception quality. To consider potential line-of-sight (LoS) and non-line-of-sight (NLoS) for low altitude flying drones, we adopt the channel model in [8] with free-space path loss and Rayleigh fading for the *UAV-APs* uplink path loss model. The channel between the bth AP and the uth UAV can be expressed as

$$h_{u,b} = [P_{\mathrm{LoS}}^{u,b}\eta_{\mathrm{LoS}} + P_{\mathrm{NLoS}}^{u,b}\eta_{\mathrm{NLoS}}](\frac{4\pi d_{u,b} f_{\mathrm{c}}^{\mathrm{UL}}}{c})^{\alpha_{\mathrm{UL}}}\beta_{u,b}, \quad (1)$$

where $P_{\mathrm{LoS}}^{u,b}$ is given by [8].

2) AP-VR Uplink

We consider Rayleigh fading for multi-input-single-output (MISO) transmission between AP group and VR user [5]. The channel between the bth AP and vth VR user is denoted as $h_{b,v} = d_{b,v}^{-\alpha_{\mathrm{DL}}}\beta_{b,v}$, where $d_{b,v}$ represents the distance between bth AP and the vth VR user, α_{DL} represents the AP-VR uplink path loss exponent, and $\beta_{b,v}$ denotes the Rayleigh small-scale fading.

B. User Correlation Model and Video Decoding Model

To facilitate effective broadcasting of video resources, it is necessary to split the captured video resource into small tiles, which can be decoded individually. We define that each tile contains color information for 30° × 30° square in 3D global space. The size of one tile is defined as μM_{T} bits, where μ is the compression rate. We also denote the overall tile set as $\mathcal{J}$, which is provided by the set of UAVs $\mathcal{U}$. The tiles set generated at time t is denoted as $\mathcal{J}_t$. Each UAV provides 6×12 tiles for 180° × 360° full-view. Since the field-of-view (FoV) of human is defined as 210° × 150° [4]. Thus, we have 5×7 tiles in one users' request set $\mathcal{J}_t^v,\ v \in \mathcal{V}$ at time t, i.e. $|\mathcal{J}_t^v| = 21$. All VR users' FoV and corresponding tile requests are randomly generated within the UAV's viewpoint whose required tiles follow the rule of 3D-2D projection. By dividing the VR users' requested video frame into tiles, the VR user group $\mathcal{V}_j$, who requests the same tile j from the uth UAV, can be served via the

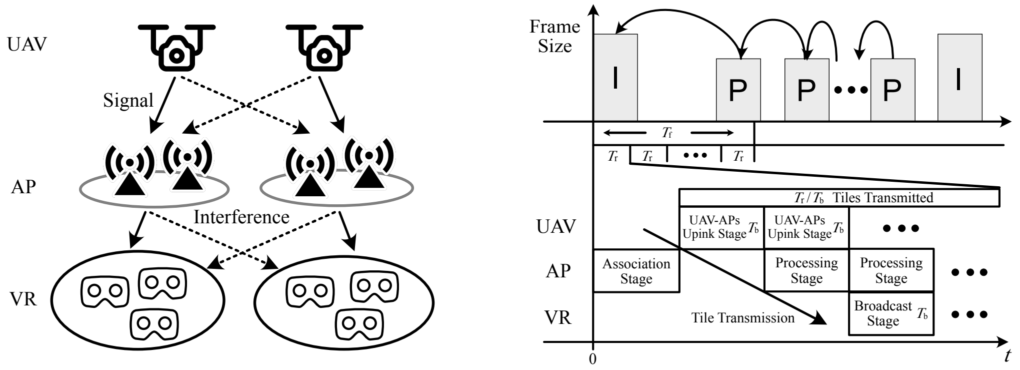

Fig. 1. Communication stages and video tiles decoding relationship for considering multi-group-multi-cast system.

broadcast channel at the same time. This highlights the potential benefit of broadcasting overlapping tiles.

As shown in Fig. 1, a set of new tiles from video frames with corresponding J_t tiles are generated every T_{f} time, i.e. at frame rate $1/T_{\mathrm{f}}$. We assume that all video resources are captured and encoded in the same frame rate and aligned in time. Considering a dependent frame decoding scheme, frames are encoded incrementally within group-of-pictures (GOP) to reduce the overall data rate. In each GOP, the first intro-coded tile (I) can be decoded independently. The predicted-coded (P) tile can be successfully decoded only when the previous tiles are successfully decoded, whose set is denoted as $\mathbf{J}_t^v$ in the vth VR user at time t.

C. Network Transmission Procedure

In our considered CF-MB network, the network performs as a DF relay system to support the tile j transmission from uth UAV to VR user group $\mathcal{V}_j$ via APs group $\mathcal{B}_t^u$. With time division duplex, the tiles with highest popularity in each UAV is firstly selected for transmission. As such, the target VR users whose requested tiles are scheduled in J_t forms the VR user group $\mathcal{V}_j$ $(j \in J_t)$. As shown in Fig. 1, this allows for the spatial reuse of frequency resources, where multiple UAVs can be served by different virtual cells at the same time. In the *UAV-APs* uplink stage, the uth UAV transmits the scheduled tile j to it associated AP group $\mathcal{B}_t^u$, which jointly receives the signal. In the processing stage, the tile is processed at the central server, whose delay is considered as a constant value and ignored in our analysis. In the broadcast stage, the APs in virtual cell $\mathcal{B}_t^u$ jointly broadcast the tile to the VR user group requesting tile j, i.e. $\mathcal{V}_j$. These stages repeated until the end of T_{r} using the same association decision.

D. Tile Transmission Model

From the perspective of CF-MB network, each AP is equipped with two antenna for full-duplex transmission where one antenna for *UAV-APs* uplink, one for *APs-VR* downlink. We assume that all UAV and VR users are equipped with one antenna. The tile transmission model can be seen as a DF relay system with linear MRC/MRT, where *UAV-APs* uplink is SIMO transmission and *APs-VR* downlink is MISO transmission.

1) The Transmission of UAV-APs Uplink

For the SIMO transmission of *UAV-APs* uplink from single UAV to multiple cooperative APs, the received signal $\gamma_{u^*,\mathcal{B}_t^u}$ at the central sever from scheduled the u^*th UAV within associated APs group $\mathcal{B}_t^{u^*}$ at time t is

$$y_{u^*,\mathcal{B}_t^u} = \sum_{b\in\mathcal{B}_t^u} w_b \Big[h_{u^*,b} s_{u^*} + \sum_{u'\in\mathcal{U}_t\backslash u^*}^{\mathcal{U}} h_{u',b} s_{u'} + n_0\Big], \quad (2)$$

where $h_{u,b}$ denotes the channel vector from the uth UAV to the bth AP, $\mathcal{U}_t$ is the current scheduled UAV, $h_{u',b}$ is the interference channel from other interfering UAVs, s_u is the signal transmitted by the uth UAV, $N_0 \sim \mathcal{CN}(0, I_N)$ represents the Gaussian white noise, w_b is a general weighted MRC scheme with weight $w_b = h_{u^*,b}^H/||\mathbf{h}_{u^*,\mathcal{B}_t^u}||_{\mathrm{F}}$, $b \in \mathcal{B}_t^u$, $||\cdot||_{\mathrm{F}}$ represents Frobenius norm, and $\mathbf{h}_{u^*,\mathcal{B}_t^u} = [h_{u^*,b_0}, ..., h_{u^*,b_{|\mathcal{B}_t^u|}}]$ is a $|\mathcal{B}_t^u| \times 1$ channel vector from target the u^*th UAV to a corresponding APs group $\mathcal{B}_t^u$.

Thus, the received SINR for tile upload from the uth UAV to access point group $\mathcal{B}_t^k$ at time t is expressed as

$$\gamma_{u^*,\mathcal{B}_t^u} = \frac{\sum\limits_{b\in\mathcal{B}_t^u} p_u |w_b h_{u,b}|^2}{\sum\limits_{b\in\mathcal{B}_t^u} \sum\limits_{u'\in\mathcal{U}\backslash u}^{\mathcal{U}} p_{u'} |w_b h_{u',b}|^2 + \sum\limits_{b\in\mathcal{B}_t^u} |w_b|^2\sigma^2}, \quad (3)$$

where p_u is the transmitting power of UAV, σ is the power of Gaussian noise. Due to the flat-fading in each broadcast slot, the received data capacity $D_{u^*,\mathcal{B}_t^{u^*}}(t)$ during resource block at the group of APs $\mathcal{B}_t^{u^*}$ from the u^*th UAV is given by $D_{u^*,\mathcal{B}_t^{u^*}} = T_{\mathrm{b}} B_{\mathrm{UL}} \log_2(1 + \gamma_{u^*,\mathcal{B}_t^{u^*}})$.

2) Tile Transmission of the APs-VR Uplink

In *APs-VR* uplink, the APs form virtual-cells to jointly broadcast the tiles to corresponding VR user groups and enhance the broadcasting quality. We adopt linear sum maximum precoding to ensure the broadcast performance [9]. With perfect channel state information (CSI), the precoding matrix in the bth AP can be given by $w_b = \alpha_b \sum_v^{\mathcal{V}_t^k} h_{b,v}^H / ||h_{b,v}||^2, \ b \in \mathcal{B}_t^k,$, where α_b is the normalize factor to ensure $||w_b||_{\mathrm{F}}^2 = 1$. Then, the signal received at the selected v^*th VR user ($v^* \in \mathcal{V}_t^u$) from the bth AP can be expressed as

$$y_{\mathcal{B}_t^u,v^*} = \sum_{b \in \mathcal{B}_t^u} h_{b,v^*} w_b s_b + \sum_{b' \in \mathcal{B} \backslash \mathcal{B}_t^u}^{\mathcal{B}} h_{b',v^*} w_{b'} s_{b'} + n_{v^*} \tag{4}$$

where w_b denotes the precoding matrix for the bth AP in group $\mathcal{B}_t^u$ at time t, and $h_{b,v}$ is the path loss for the channel between the bth AP and the vth VR user.

Based on (4), the SINR from the bth AP in APs group $\mathcal{B}_t^u$ to v^*th VR user in user group $\mathcal{V}_t^u$ at time t can be expressed as

$$\gamma_{\mathcal{B}_t^u,v^*} = \sum_{b \in \mathcal{B}_t^u} p_b |h_{b,v^*} w_b|^2 \Big(\sum_{b' \in \mathcal{B} \backslash \mathcal{B}_t^u}^{\mathcal{B}} p_{b'} |h_{b',v^*} w_{b'}|^2 + \sigma^2 \Big) \tag{5}$$

where p_b is the transmitting power of AP, σ is the power of Gaussian noise. Under given SINR, the received data $D_{\mathcal{B}_t^u,v}$ in one broadcast slot T_{b} from the APs group $\mathcal{B}_t^u$ to v^*th VR user is calculated via minimum ergodic rate in the broadcast group $D_{\mathcal{B}_t^u,v} = T_{\mathrm{b}} B_{DL} \log_2(1 + \gamma_{\mathcal{B}_t^u,v^*})$.

3) Overall Capacity

From the whole system point of view, the tiles are delivered to the requesting VR user group via the aforementioned DF network transmission. As the success of tile transmission will only occur when both the *UAV-APs* uplink and *APs-VR* downlink success. The successful transmission of jth tile can be written as the combination of successful transmission in *UAV-APs* uplink and *APs-VR* downlink as

$$\mathbb{1}[D_{u,v} \geq \mu M_{\mathrm{T}}] = \mathbb{1}[D_{u,\mathcal{B}_t^u} \geq \mu M_{\mathrm{T}}] \wedge \mathbb{1}[D_{\mathcal{B}_t^u,v} \geq \mu M_{\mathrm{T}}], \tag{6}$$

where μM_{T} is the size of tile to be transmitted, $\mathbb{1}[x] = 1$ as x is true, $\mathbb{1}[x] = 0$, otherwise. $\wedge$ is logical and operation. $\mathbb{1}[x] \wedge \mathbb{1}[y] = 1$ as x and y is true.

E. Quality-of-experience Metric for VR Users

Generally, for video-based VR service, it is common to define the QoE as the break-in-presence (BIP), which is tightly correlated to the video quality. It is common to measure the received amount of information in tiles or frames via the Peak Signal-to-Noise (PSNR) value. To model the QoE with PSNR value, we jointly consider the transmission and decoding of tiles. We first define the successful decoded tiles at the vth VR user as set $\mathbf{J}_t^v$ at time t. We then refactorize the original PSNR function, which measures the pixel-level information amount, to that of tile-level. By doing so, we can quantify the decoded QoE with PSNR value inside the vth VR user field-of-view at time t using viewport-PSNR (V-PSNR) [10] as

$$\text{V-PSNR}_t^v = 10 \log_{10}\Big(\frac{1}{1 + \frac{1}{|\mathcal{J}_t^v|}(|\mathcal{J}_t^v| - \sum_{j \in \mathcal{J}_t^v} \mathbb{1}[j \in \mathbf{J}_t^v])} \Big), \tag{7}$$

where $\mathbf{J}_t^v$ is the decoded tile set, and $\mathbf{J}_t^v$ represents the actual decoded tiles at time t in the vth VR user ($\mathbf{J}_t^v \subseteq \mathcal{J}_t^v$). In (7), $\sum_{j \in \mathcal{J}_t^v} \mathbb{1}[j \in \mathbf{J}_t^v]$ denotes the number of successfully decoded tiles in vth VR user at time t. The V-PSNR value gives $10 \log_{10} |\mathcal{J}_t^v|$ if all the tiles requested by the vth VR user are transmitted successfully. The tile j can be decoded if its and its dependent tiles are successfully transmitted or decoded

$$\mathbb{1}[j \in \mathbf{J}_t^v] = \begin{cases} \mathbb{1}[D_{u,v} \geq \mu M_{\mathrm{T}}], & t < T_{\mathrm{f}}, \\ \mathbb{1}[D_{u,v} \geq \mu M_{\mathrm{T}}] \wedge \mathbb{1}[j' \in \mathbf{J}_t^v], & t \geq T_{\mathrm{f}} \end{cases} \tag{8}$$

where $\mathbb{1}[D_{u,v} \geq \mu M_{\mathrm{T}}]$ is given in (6), j and j' are dependent tiles, j is required to be decoded with j' incrementally, i.e. j depends j' to decode $j \to j'$. In each GOP, when $t < T_{\mathrm{f}}$, the tile is the first one which can be decoded independently.

F. Problem Formulation

We aim to study how the CF-MB network supports the tile transmission from UAVs to VR users and enhance the QoE of VR users by dynamically adjusting the association decisions. In each broadcast slot, the network generates an state $S_{t_{\mathrm{b}}}$, which indicates VR users' request, UAVs' position, VR users' V-PSNR, *UAV-APs* uplink's, *APs-UAV downlink*'s channel information, and etc. The network state $S_{t_{\mathrm{b}}+1}$ in next broadcast slot is jointly decided by the current system state $S_{t_{\mathrm{b}}}$ and association decision. such that it forms Markov decision process (MDP) problem. In this problem, we denote the association policy π_{a} as the distribution mapping from the current environment state to the selection of each UAV and corresponding VR user group. Thus, our optimization target can be defined as maximizing the accumulative V-PSNR gain over broadcast slots in T_{GOP}

via finding the optimal π_{s} and π_{a}

$$\max_{\pi_{\mathrm{a}}} \mathbb{E}[\sum_{t_{\mathrm{b}}=0}^{T_{\mathrm{GOP}}} \underbrace{\sum_{j \in J_{t_{\mathrm{b}}}} \sum_{v \in \mathcal{V}_j} \Delta \text{V-PSNR}_{t_{\mathrm{b}}}^{v}]}_{\text{V-PSNR Gain in } T_{\mathrm{b}} \text{ for scheduled tile set } J_{t_{\mathrm{b}}}}, \quad (9)$$

where the V-PSNR gain is denoted as $\Delta\text{V-PSNR}_{t_{\mathrm{b}}}^{v} = \text{V-PSNR}_{t_{\mathrm{b}}}^{v} - \text{V-PSNR}_{t_{\mathrm{b}}-1}^{v}$.

G. Conventional Approaches

We adopt two conventional network schemes to handle the association problem, which are cell-based (CB) and cell-free MIMO (CF) associations: 1) In CB network, each AP is an individual cell, where each AP makes its decision based on its observation independently cooperation. This scheme may bring high inter-cell interference and poor cell-edge performance; and 2) In CF network, all APs cooperatively receive one tile in every *UAV-APs* uplink stage and broadcast one tile in the broadcast stage. In another word, all APs are grouped in one virtual cell. In this scheme, the tile with the highest priority among all tiles in all UAVs is selected to be transmitted.

III. Reinforcement Learning Approach for Association

With the geometry-correlated VR users' request, which provides another degree-of-freedom in system design, it calls for an intelligent association algorithm to spatially reuse the frequency resource by grouping APs into several virtual cells, while the large number of APs introduce high dimensional state-action space. To address this issue, we divide the overall association optimization target (9) into effective and non-effective areas for each AP. It should be noted that the observation range of each AP is limited due to the fluctuated nature of the wireless signal. In this way, we can formulate our association problem as a networked decentralized partially observable Markov decision processes (ND-POMDP) problem [11], which is a factored version of Decentralized-POMDP problem with mean field theorem [12]. The state space is denoted as $\mathcal{S}$ ($s \in \mathcal{S}$). The joint action space can be denoted as $\mathcal{A} = \prod_{b \in \mathcal{B}^b} \mathcal{A}_b$, where $\mathcal{A}_b$ is the set of local action space of the bth AP. Then, the value function for the bth AP in state s ($s \in \mathcal{S}$) can be written as the function of joint action space of current AP and its neighbours

$$v_b(s) = \sum_{a_b \in \mathcal{A}_b} \pi_{\mathrm{a}}^{b}(a_b|s, (\mathbf{a}_{-b}))\mathbb{E}_{a_b,(\mathbf{a}_{-b})\sim(\pi_{\mathrm{a}}^{-b})}[q_b(s, a_b, (\mathbf{a}_{-b}))]. \quad (10)$$

where $\mathcal{A}_b$ is the action set of bth AP, $\mathbf{a}_{-b}$ present the joint action for bth AP's neighbors, $\pi_{\mathrm{a}}^{-b} = \prod_{a_j \in \mathbf{a}_{-b}} \pi_{\mathrm{a}}^{b'}(a_j|s)$ is the joint policy for AP's neighbors, $\pi_{\mathrm{a}}^{b'}$ is the association policy for b'th AP ($b' \in \mathcal{B}^b/b$). The Q function for bth AP is denoted as q_b. As such, the size of the problem is largely reduced by factorizing the problem as local sub-problem of each AP and its neighbors and treating the rest of APs as part of environment.

Algorithm 1: Distributed DRL based association.

```
Initiate environment $Env$, state $S_0$, and association network parameters.
repeat
    if Game end then
        Reset $Env$ and $t = 0$, obtain new $S_0$, $O_0^{\mathrm{s}}$
    Schedule tiles $J_t$ based on popularity
    for $b \in \mathcal{B}$ do
        Obtain $O_t^b$ from $S_t$, $S_{t-1}$ and scheduled tiles $J_t$
        Select an action $A_t^b$ with (11)
    Tiles are transmitted from $u$th UAV to corresponding VR users via APs group $\mathcal{B}_t^u$, and obtain $S_{t+1}$
    for $b \in \mathcal{B}$ do
        Calculate reward $R_t^b$ in effective area
        Push tuple $(O_t^b, A_t^b, R_t^b)$ to experience replay
        Train and update association network's parameters.
    if $t$ can be divided by $T_{federated}$ then
        Perform FedAvg among APs $\mathcal{B}$
until Converge
```

Then, we propose a multi-agent DRL approach to solve the proposing ND-POMDP problem with the idea of the mean-field theorem [12], where one AP is associated with one agent. To capture the geometry information, we design a fuzzy version of the state based on a grid-based observation, where the geometry correlated information is mapped into grids. As shown in Fig. 2, for each UAV, three grid-maps correspond to the position of UAVs, APs, and the VR user group requesting the currently scheduled tiles ($\mathcal{V}_j, j \in J_t$). The value in each UAVs and APs grid maps is 1 if the node exists in that grid. For the VR user group grid-map, the value in each grid is the summation number of tile requests from the VR users in that grid, which is normalized over the maximum number of tiles' requests in each grid.

To capture the spatial information in grid observations among UAV, AP, and VR users, we introduce convolutional layers to encode the observation into following neural layers. The benefit of applying convolutional layers in communication problems has been shown by previous researches [7]. The convolutional layers can learn to estimate the potential signal and interference. As shown in Fig. 2, we design five layers of convolu-

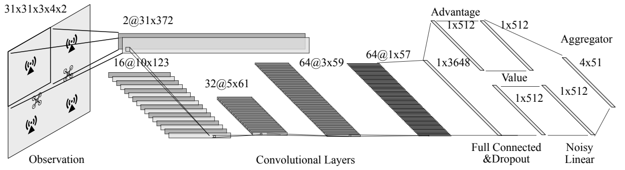

Fig. 2. Network Structure of Distributed Association Agent.

tional layers and one linear layer to jointly encode the observation into a hidden vector. The hidden vector is then processed via a rainbow agent to generate the final action. In our problem, the main motivation for applying a rainbow agent is its benefits from distributional DRL [13]. Due to the partial observability and random nature of wireless network, the value of state s is not a static number. With distributional DRL, the distribution of value function is directly estimated in distribution form. The value estimation generated by the neural network for each action is a discrete mapping from the actual continues value distribution to a distributive value vector with N_{atom} atoms. The network is trained by minimising the Kullbeck-Leibler divergence between the estimated distribution and the target distribution. The algorithm is presented in **Algorithm 1**.

At the time t, each AP observes its local observation from state and select an action A_t^b based on the value distribution estimation from the neural network. We adopt the Boltzmann policy to capture actions with relatively small return, which potentially benefit the overall environment via effective cooperation. The Boltzmann policy for bth AP in state s can be formulated as [12]

$$\pi_{\text{a}}^{b}(a_b|s,(\mathbf{a}_{-b}))=\frac{\exp-\beta q_b(s,a_b,(\mathbf{a}_{-b}))}{\sum_{a_b\in\mathcal{A}_b}\exp-\beta q_b(s,a_b,(\mathbf{a}_{-b}))},\tag{11}$$

where β is the temperature for Boltzmann policy, $q_b(s,a_b,(\mathbf{a}_{-b}))$ is the Q function value estimated by the neural network. Together with all other APs, after serving UAV and corresponding VR users cooperatively, the bth AP receives the reward $R_t^b=\sum_{v\in\mathcal{V}^b}\Delta\text{V-PSNR}_t^v$, where $\mathcal{V}^b$ is the VR users in bth AP's observation range.

For our considered cooperation multi-agent system, it has been shown that communication among agents can substantially improve its performance. We apply FL via federated average (FedAvg) algorithm to combine useful knowledge from all other APs. The FedAvg algorithm performs averaging every $T_{\text{federated}}$ time intervals, which is naturally suitable for our network with a central server. Besides, with FL, all agents can be seen as fully cooperative agents with the same optimization target (V-PSNR gain) and global learning model ($\pi_b=\pi_{b'},b,b'\in\mathcal{B}$). It has been shown that our considered problem can be directly solved using MDP methods with joint action space $\mathcal{A}=\prod_{b\in\mathcal{B}}\mathcal{A}_b$ of fully cooperative agents and global updated policy in each agent. This guarantees the convergence of our proposing multi-agent algorithm.

IV. Simulation Result

In the simulation, we set the number of VR users as $|\mathcal{V}|=120$, the VR users are distributed following PCP, whose cluster radius is set as $r_{\text{c}}=20m$, the number of UAVs is $|\mathcal{U}|=4$. We set the number of AP as $|\mathcal{B}|=9$, which are located in a 3×3 grid with $30\,\text{m}$ gap inside the serving area which is $80\,\text{m}\times80\,\text{m}$ square. Each AP can observe $60m\times60m$ squared effective area surrounding itself. The time period of learning algorithms contains $10T_{\text{b}}$, which means the association policy is updated after broadcasting 10 tiles. The AP's and UAV's effective isotropic radiated power (EIRP) is set as $48\,\text{dBm}$. The noise power is $-91\,\text{dBm}$. The *UAV-APs* uplink and *APs-VR* downlink channel's bandwidth are $5\,\text{MHz}$, the center frequency is $4.5\,\text{GHz}$ and $5.5\,\text{GHz}$, and the path-loss parameter is $\alpha_{\text{UL}}=2$ and $\alpha_{\text{DL}}=4$, respectively.

With CF and CB as baselines, we further introduce the centralized DRL algorithm to show the effectiveness of our proposed algorithm. To ease the presentation of V-PSNR, we normalize the resulting V-PSNR value into $[0,5]$ (5 frames in each GOP). To show the risk of our algorithm, we use a standard derivative (SD) error bar to show the performance. In the following, we use "Centralized" and "Distributed DRL" to denote the centralized DRL association algorithm and federated distributed DRL algorithm, respectively.

Fig. 3 plots the V-PSNR value versus the number of VR users. We observe that all algorithms' V-PSNR stay nearly unchanged with increasing numbers of VR users in CF-MB network. This matches our expectation for CF-MB network, where the UAV-APs cooperative reception enhance the received signal from UAV and the APs-VR broadcasting is not sensitive to the number of receiving VR users. It is worth to mention that we only train a single model using random VR users and obtain similar results with different numbers of VR users setting, showing the scalability of proposed algorithms.

Fig. 4 plots the V-PSNR versus the number of broadcast slots, which also reveals the slot utilization of our proposed algorithms. Remind that in our considered environment, 4 UAV holds 288 tiles in total. If we set large T_{b} and fewer broadcast slots, then two tiles should be fully transmitted successfully within one broadcast slot (160 slots). If we set small T_{b} and more broadcast slots,

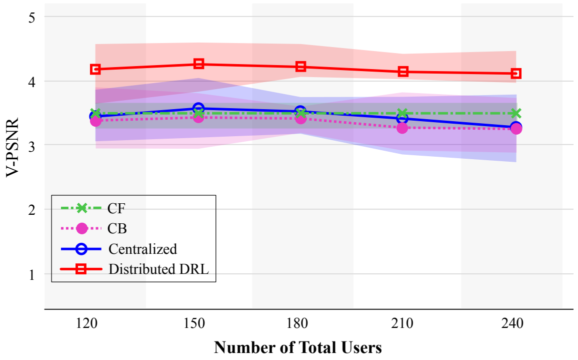

Fig. 3. V-PSNR of our proposing algorithms with different number of VR users.

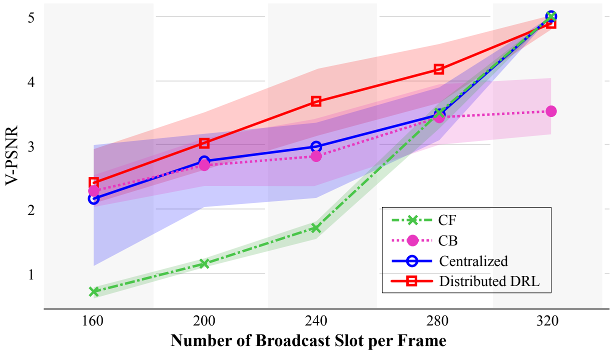

Fig. 4. V-PSNR of our proposing algorithms with different broadcast slots in each frame duration.

each tile can occupy one broadcast slot individually (320 slots). We observe that the V-PSNR of CF association method increases with the number of broadcast slots, as the chances of transmission increase. We also observe that the V-PSNR value of CB association approach decrease with the increasing of broadcast slots. The reason is that high inter-cell interference results in low cell-edge performance. We highlight that the V-PSNR value of learning-based algorithms is higher than CB and CF. And the Distributed DRL outperform the Centralized approach with all different number of broadcast slots with small SD.

V. Conclusion

In this paper, we introduced a cell-free multi-group broadcast network for real-time VR video transmission from UAVs to VR users for experience enhancement in a sports event. To optimise the quality-of-experience of VR users with dependent decoded video resources and correlated VR users, we highlighted the importance of the dynamical association of APs. With popularity-based scheduler, we first introduced cell-based and cell-free association algorithms to solve each sub-problem individually. To explore the learning-based dynamic association algorithm, we then propose a multi-agent deep reinforcement learning algorithm, which captures the environment via convolutional layers. Our results demonstrated that both distributed APs with federated learning algorithms can effectively handle a large number of APs and VR users and outperform the centralized algorithm and non-learning-based approach with different environment settings.

References

[1] F. Hu, Y. Deng, W. Saad, M. Bennis, and A. H. Aghvami, “Cellular-Connected Wireless Virtual Reality: Requirements, Challenges, and Solutions,” *IEEE Commun. Mag.*, vol. 58, pp. 105–111, May 2020.

[2] M. Chen, W. Saad, C. Yin, and M. Debbah, “Data Correlation-Aware Resource Management in Wireless Virtual Reality (VR): An Echo State Transfer Learning Approach,” *IEEE Trans. Commun.*, Feb. 2019.

[3] C. Perfecto, M. S. Elbamby, J. D. Ser, and M. Bennis, “Taming the Latency in Multi-user VR 360: A QoE-aware Deep Learning-aided Multicast Framework,” *IEEE Trans. Commun.*, vol. 68, no. 4, Apr. 2020. [Online]. Available: http://arxiv.org/abs/1811.07388

[4] 3GPP, “3GPP TS 26.247 Progressive Download and Dynamic Adaptive Streaming over HTTP (3GP-DASH),” 2018.

[5] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, “Cell-Free Massive MIMO Versus Small Cells,” *IEEE Trans. Wirel. Commun.*, vol. 16, no. 3, pp. 1834–1850, Jan. 2017.

[6] S. Buzzi and C. D’Andrea, “Cell-free massive MIMO: User-centric approach,” *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 706–709, Dec. 2017.

[7] W. Cui, K. Shen, and W. Yu, “Spatial Deep Learning for Wireless Scheduling,” *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1248–1261, Jun. 2019.

[8] M. Mozaffari, W. Saad, M. Bennis, and M. Debbah, “Unmanned Aerial Vehicle with Underlaid Device-to-Device Communications: Performance and Tradeoffs,” *IEEE Trans. Wirel. Commun.*, vol. 15, no. 6, pp. 3949–3963, Feb. 2016.

[9] J. Joung, H. D. Nguyen, P. H. Tan, and S. Sun, “Multicast linear precoding for MIMO-OFDM systems,” *IEEE Commun. Lett.*, vol. 19, no. 6, pp. 993–996, Apr. 2015.

[10] C. Li, M. Xu, L. Jiang, S. Zhang, and X. Tao, “Viewport Proposal CNN for 360° Video Quality Assessment,” in *IEEE Conf. Comput. Vis. Pattern Recognit.*, Long Beach, CA, Jun. 2019.

[11] N. Ranjit, V. Pradeep, T. Milind, and Y. Makoto, “Networked distributed POMDPs: a synthesis of distributed constraint optimization and POMDPs,” in *AAAI’05*, Pittsburgh, Pennsylvania, USA, 2005, pp. 133–139. [Online]. Available: https://www.aaai.org/Papers/AAAI/2005/AAAI05-022.pdf

[12] Y. Yang, R. Luo, M. Li, M. Zhou, W. Zhang, and J. Wang, “Mean Field Multi-Agent Reinforcement Learning,” *arXiv preprint arXiv:1802.05438*, Feb. 2018. [Online]. Available: http://arxiv.org/abs/1802.05438

[13] M. Hessel, J. Modayil, H. Van Hasselt, T. Schaul, G. Ostrovski, W. Dabney, D. Horgan, B. Piot, M. Azar, and D. Silver, “Rainbow: Combining Improvements in Deep Reinforcement Learning,” in *32nd AAAI 2018*. AAAI press, Oct. 2018, pp. 3215–3222.