QoE-Aware Power Control for UAV-aided Media Transmission with Reinforcement Learning

Yuzhen Ding*, Donghua Jiang*, Jinhao Huang*, Liang Xiao*†, Sicong Liu*, Yuliang Tang*, Huaiyu Dai‡

*Dept. Communication Engineering, Xiamen University, Xiamen, China. Email: liusc@xmu.edu.cn

†National Mobile Communications Research Laboratory, Southeast University, Nanjing, China.

‡Dept. of Electrical and Computer Engineering, NC state University, Raleigh, NC.

***Abstract*—Unmanned aerial vehicles (UAVs) are widely utilized to capture and compress videos of the target area and then transmit the processed videos to the control station (CS) on the ground. The media transmissions in the UAV-aided network face many challenges due to the highly dynamic network topology and limited resources such as bandwidth and energy. This paper introduces a media transmission scheme in the UAV-aided network utilizing reinforcement learning algorithms to efficiently process and transmit the captured video, which is able to improve the quality-of-experience (QoE) and reduce the energy consumption. Exploiting the proposed reinforcement learning algorithm, the UAV dynamically selects the quantization parameter in the source coding process and determines the transmit power without knowing the video transmission model. Simulation results demonstrate that the proposed scheme is capable of achieving a higher video quality and utility with lower energy consumption compared with the state-of-the-art schemes.**

***Index Terms*—Unmanned aerial vehicle, video coding, power control, reinforcement learning.**

I. Introduction

Unmanned aerial vehicles (UAVs) equipped with camera sensors are widely deployed in many applications, such as area surveillance, searching tasks and border patrol, to capture videos of the target area and transmit them to the control station (CS) [1], [2]. Embedded with processors and storage devices, UAVs with enhanced processing ability, high mobility and low energy consumption can efficiently process and transmit the video data in real time. However, it is hard to realize the video transmission with high quality-of-experience (QoE) owing to the high dynamism of the UAV-aided scenario, the enormous size of the captured video and the limited wireless transmission resources [3].

For the sake of reducing the enormous size of the video data and relieving the pressure of processing and transmission on the UAV with limited storage and computational capability, video compression coding techniques, such as H.262, H.263, H.264/MPEG-4 Advanced Video Coding and H.265/High Efficiency Video Coding, can be exploited [4]. The quantization parameter (QP), as a key parameter in the video compression process, affects the compression quality and processing delay [5]. Conventional media coding parameter selection schemes usually use constant QP [6] or choose the QP based on a known video transmission model and network impairment model [7]. However, in the dynamic UAV-aided networks, it is difficult for the UAV to obtain the specific channel model and video transmission model, so the UAV cannot adjust the QP dynamically to improve the video compression performance. In the process of video transmission, the selection of the transmit power of the UAV affects the signal-to-noise ratio (SNR) and the energy consumption. However, many existing schemes, which usually set the transmit power fixed during the media transmission [8], or reduce it to minimize the energy consumption upon user requirements [9], rarely realize a dynamic power control. The video quality, end-to-end delay and energy consumption are three significant metrics to measure the performance in the process of video compression and transmission, which is a trade-off in practice to improve the video QoE. It is difficult for the state-of-the-art schemes to jointly determine a suitable QP and transmit power that can achieve high video QoE and low energy consumption simultaneously in the highly dynamic UAV-aided networks.

In order to improve the video QoE and reduce the energy consumption, a UAV-aided media transmission sheme is proposed in this paper, where the UAV selects the QP to compress the video using the H.264 standard and chooses the transmit power to transmit the processed video to the CS. It is a challenge for the UAV to select the optimal transmission policy, including the QP and transmit power, without knowing the specific video transmission model. Since the current decision of transmission policy only depends on the current state rather than the past ones, a Markov decision process (MDP) can be used to model the video transmission scenario [10]. Therefore, the emerging powerful reinforcement learning (RL) algorithms are suitable for selection of the QP and transmit power to improve the video QoE and utility

This work was supported in part by the National Natural Science Foundation of China under Grants 61671396, 61971366, 61901403, 91638204, 61371081 and 61871339, and the open research fund of National Mobile Communications Research Laboratory, Southeast University (No. 2018D08), and the Natural Science Foundation of Fujian Province of China under Grants 2019J01843 and 2019J05001, and the Fundamental Research Funds for the Central Universities of China (No. 20720190029 and No. 20720190034), and the US National Science Foundation under grants EARS-1444009 and CNS-1824518. (*Corresponding author: Sicong Liu*).

978-1-7281-0962-6/19/$31.00 ©2019 IEEE

and reduce the energy consumption via trial and error. More specifically, upon receiving a task request with the event priority and the available channel bandwidth from the CS, the UAV observes its distance to the CS and then dynamically selects the media transmission policy based on RL algorithms. Simulation results demonstrate the capability of our proposed scheme to significantly improve the UAV-aided media transmission performance consisting of the video quality and energy consumption compared with the current state of the art, such as the logarithmic selection method (LSM) based scheme [7] and the greedy-based power allocation (GPA) algorithm [11].

The main contributions of this work are summarized as below:

- We propose a UAV-aided media transmission scheme, in which the UAV compresses the captured video with the selected QP and transmits the compressed video with the chosen transmit power to the CS on the ground to improve the video QoE and reduce the energy consumption.
- The RL technique is applied in the proposed media transmission scheme to select the optimal QP and transmit power without knowing the specific video transmission model.
- We prove that the RL-based scheme can achieve the optimal performance in the dynamic media transmission process, and the performance bound regarding the video quality, energy consumption and utility is provided.

The rest of this paper is organized as the following. Section II discusses the related works. The UAV-aided network model and the video transmission model are introduced in Section III. In Section IV, we describe the proposed RL-based UAV-aided media transmission scheme. The performance analysis is derived in Section V. Section VI reports simulation results and Section VII concludes this paper.

II. Related Works

There have been extensive research works on improving the performance of video compression coding. For instance, a global motion assisted video encoding method proposed in [12] investigates a bit allocation scheme to decrease the coding complexity. The forward error correction coding technique is proposed in [13] to improve the video delivery quality. A motion estimation algorithm is proposed in [14] to efficiently reduce the computational complexity and shorten the encoding time of video coding.

Multiple studies have been conducted for coding parameter selection in video compression. A parameter optimization framework for the H.264 encoder investigated in [15] applies function fitting to achieve the QP with the optimal trade-off among bit rate, quality of multimedia contents, processing delay and power consumption. A QP selection scheme developed in [7] chooses the QP related to the total retransmission bits and the logarithmic relationship therein. A parametric domain rate control algorithm proposed in [16] selects the value of QP through a Lagrange multiplier.

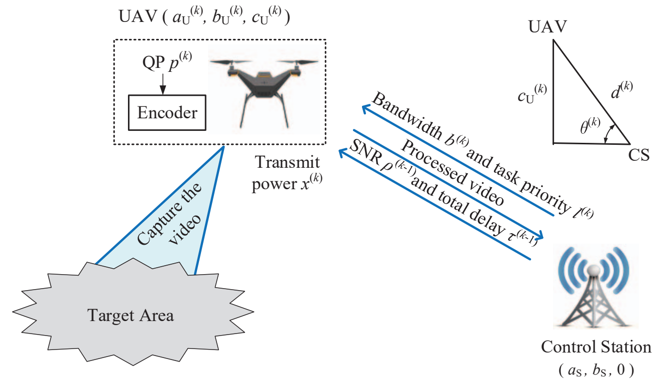


Fig. 1: Media transmission process in a UAV-aided network, in which the UAV receives the feedback information and the priority upon the request from the CS and sends the processed video of the target area to the CS.

Numerous works on transmit power allocation in UAV-aided media transmission have been investigated. For example, a transmit power management scheme applies a tree structure optimization method to minimize the transmission energy under acceptable video quality [17]. A multi-UAV enabled wireless communication system applies a successive convex optimization based iterative algorithm to jointly optimize the transmit power and trajectory [8]. The UAV-supported ultra dense networks proposed in [18] utilize the dynamic game theory among all the UAVs to realize the distributed optimization for power control and improve the energy efficiency.

Recently, RL techniques have been introduced to video coding, media transmission and UAV-aided networks. A RL-based control system adaptively assesses the video quality and calculates the code rate according to the standard network information [19]. An automatic navigation scheme for UAVs proposed in [20] uses only the sensory information and the GPS signal of the local environment to implement autonomous navigation and obstacle avoidance. An MDP-based video adaptation scheme applies value iteration and Q learning to determine the quality of the next chunk, which can reduce the occurrence of video freezing or re-buffering [21].

III. System Model

A. Network Model

The UAV-aided network model for media transmission is shown in Fig. 1, where a UAV located at the coordinate of $L_{\mathrm{U}}^{(k)} = (a_{\mathrm{U}}^{(k)}, b_{\mathrm{U}}^{(k)}, c_{\mathrm{U}}^{(k)})$ with flexible mobility and a good channel state captures the video of the target area at time slot k and transmits it to the CS. The distance and the elevation angle between the UAV and the CS are denoted by $d^{(k)}$ and $\theta^{(k)}$, respectively. More specifically, the UAV compresses the captured video with the selected QP $p^{(k)} \in [P_1, P_2]$ and transmits it using the selected transmit power $x^{(k)} \in [X_1, X_2]$ to tackle the bottleneck of limited storage and reduce the energy consumption. The CS located at the coordinate of

$L_{\mathrm{C}} = (a_{\mathrm{S}}, b_{\mathrm{S}}, 0)$ receives the compressed video for further processing such as decoding and playing.

The CS sends a task request and the available channel bandwidth denoted by $b^{(k)}$ to the UAV to request for the specific real-time media data at time slot k. For example, when a traffic accident happens, the traffic control centre can send a request to the UAV on site to obtain the real-time media data about the accident spot. The scene the UAV is intended to capture is called the target area. If there is more than one target area that needs capturing, the CS can send a request with priority $l^{(k)}(0 \leq l^{(k)} \leq L)$, which is divided into $L+1$ levels and reflects the emergency of an event at time slot k. More specifically, a higher priority $l^{(k)}$ indicates a more urgent event, and $l^{(k)} = 0$ means that no event happens so the UAV does not need to capture videos.

Upon receiving the task request as well as the available channel bandwidth, the UAV selects a suitable QP $p^{(k)}$ to compress the captured video and a proper transmit power $x^{(k)}$ to transmit the processed video to the CS. The CS sends the feedback information including the SNR of the received signal and the end-to-end delay back to the UAV.

B. Video Transmission Model

The UAV first compresses the captured video and uses the peak signal-to-noise ratio (PSNR) denoted by $\varpi^{(k)}$, which is a commonly used metric to measure the video compression quality and the QoE. Then the UAV transmits the compressed video to the CS, which is commonly propagated in a satisfactory channel with few obstacles. The channel power gain of the UAV-CS link denoted by $g^{(k)}$ can be evaluated by the channel path loss model, such as the model utilized in [22]. The channel path loss model usually includes two components, i.e., the line-of-sight (LoS) path and non-line-of-sight (NLoS) path with a higher path loss due to the reflection and shadowing effect. The SNR of the received signal at the CS is used to measure the transmission reliability and can be given by $\rho^{(k)} = x^{(k)} g^{(k)} / \sigma_n^2$, where σ_n^2 is the power of the zero-mean additive white Gaussian noise (AWGN) at the receiver. The video transmission between the UAV and the CS will be terminated when the noise is sufficiently large and the SNR is too low.

In video transmission tasks, the end-to-end delay is also an influential factor to the overall QoE of the video users. Assuming that the queuing delay and propagation delay are sufficiently small so that they can be ignored for simplicity, then the end-to-end delay of the video transmission denoted by $\tau^{(k)}$ mainly including the processing delay, i.e., encoding time, and the transmission delay. In addition, the energy consumption of the transmission denoted by $c^{(k)}$ is an important performance metric in power control and should be considered.

For convenience of reference, the frequently used symbols are listed in Table I.

TABLE I: List of frequently used symbols.

$d^{(k)}$	distance of the UAV-CS link
$\theta^{(k)}$	elevation angle between the UAV and the CS
$p^{(k)}$	QP used in the compression
$x^{(k)}$	transmit power at time slot k
$l^{(k)}$	event priority
$b^{(k)}$	available channel bandwidth
$g^{(k)}$	channel power gain of UAV-CS link
$\rho^{(k)}$	SNR of the transmission
$\varpi^{(k)}$	PSNR of the compressed video
σ_n^2	noise power
$\tau^{(k)}$	end-to-end delay of the transmission
$c^{(k)}$	energy consumption of the transmission
$u^{(k)}$	utility of the UAV-aided media transmission

IV. Reinforcement-Learning-Based Media Transmission Scheme

As the video compression and transmission process in the UAV-aided networks can be formulated as an MDP, an RL-based UAV-aided media transmission (RUMT) scheme that applies RL algorithms such as Q-learning is proposed to overcome the difficulty of selecting the optimal media transmission policy without knowing the specific video transmission model. More specifically, the UAV observes the states of environment and selects the QP and transmit power maximizing the Q function to compress the captured video and transmit the processed video to the CS. The Q function therein is the discounted long-term reward for each state-action pair and will be updated according to the Bellman equation iteratively [23], which is exploited in the RL framework to improve the video QoE and reduce the energy consumption.

The pseudo-code of the proposed RUMT algorithm is shown in **Algorithm 1**. The UAV can select the media transmission policy consisting of the QP $p^{(k)}$ and transmit power $x^{(k)}$ according to the current transmission state via trial and error. More specifically, the UAV observes the PSNR $\varpi^{(k-1)}$ of the captured video and receives the information of the previous video transmission including the SNR $\rho^{(k-1)}$ and the end-to-end delay $\tau^{(k-1)}$ from the feedback of the CS after accomplishing a task. Upon receiving a task with priority requirement $l^{(k)}$ and the available channel bandwidth $b^{(k)}$ from the CS, the distance $d^{(k)}$ between the UAV and the CS is measured and the system state $\mathbf{s}^{(k)} = [d^{(k)}, l^{(k)}, b^{(k)}, \varpi^{(k-1)}, \rho^{(k-1)}, \tau^{(k-1)}]$ is formulated. The UAV selects the media transmission policy $\mathbf{x}^{(k)} = [p^{(k)}, x^{(k)}] \in \mathbf{X}$ based on the ϵ-greedy criterion. The probability of choosing the optimal transmission policy is given by

$$\Pr\left(\mathbf{x}^{(k)} = \mathbf{x}^*\right) = \begin{cases} 1-\epsilon, & \mathbf{x}^* = \arg\max_{\mathbf{x}} Q\left(\mathbf{s}^{(k)}, \mathbf{x}\right) \\ \frac{\epsilon}{|\mathbf{X}|-1}, & \text{otherwise.} \end{cases} \quad (1)$$

The UAV then compresses the current captured video data with the selected QP $p^{(k)}$ and measures the PSNR $\hat{\varpi}^{(k)}$ and

Algorithm 1: RL-based UAV-aided Media Transmission (RUMT) algorithm

```
1  Initialize learning rate α, discount factor δ, initial state
   s^0, initial Q function Q(s, x) = 0, initial value
   function V(s) = 0 ;
2  for k = 1, 2, ... do
3      UAV receives a request with priority l^(k) and
       channel bandwidth b^(k) from CS ;
4      Observe and formulate the current state
       s^(k) = [d^(k), l^(k), b^(k), ϖ^(k−1), ρ^(k−1), τ^(k−1)] ;
5      Choose the policy x^(k) = [p^(k), x^(k)] according to
       the ϵ-greedy criterion ;
6      Compress the captured video with the selected QP
       p^(k) ;
7      Measure the PSNR ϖ^(k) and the processing delay
       ;
8      Transmit the processed video to the CS with the
       selected transmit power x^(k) ;
9      Receive the feedback information of SNR ρ^(k) and
       total delay τ^(k) from CS ;
10     Calculate the energy consumption c^(k) of the
       transmission ;
11     Obtain the utility u^(k) via (2) ;
12     Observe the next state
       s^(k+1) = [d^(k+1), l^(k+1), b^(k+1), ϖ^(k), ρ^(k), τ^(k)] ;
13     Update Q(s^(k), x^(k)) via Q(s^(k), x^(k)) ←
       (1 − α)Q(s^(k), x^(k)) + α(u^(k) + δV(s^(k+1))) ;
14     Update V(s^(k)) via V(s^(k)) = max_x Q(s^(k), x) ;
15 end
```

the encoding time using a video encoding software such as x264 encoder. Afterwards, the processed video is transmitted to the CS with the selected transmit power $x^{(k)}$ after measuring the distance and elevation angle as shown in Fig. 1 with the help of GPS. The CS measures the SNR of the received signal $\hat{\rho}^{(k)}$ and the transmission delay to obtain the end-to-end delay $\hat{\tau}^{(k)}$, and then sends the feedback information to the UAV. The UAV calculates the energy consumption by $c^{(k)} = x^{(k)} T_{\text{tx}}$, where T_{tx} is the video transmission time. Thus the utility of the UAV denoted by $u^{(k)}$ is obtained by

$$u^{(k)} = \hat{\varpi}^{(k)} + \hat{\rho}^{(k)} - \varphi l^{(k)} \hat{\tau}^{(k)} - \xi c^{(k)}, \tag{2}$$

where φ is the effective delay coefficient used to represent the cost per unit delay, and ξ is the cost per unit energy consumption. It can be noted from (2) that the utility is dependent on the PSNR, SNR, end-to-end delay and energy consumption. The higher the PSNR and SNR, the better the utility. On the other hand, the delay and energy consumption will play a negative role on the total utility, where the level of event priority will further determine the extent to which the delay deteriorates the utility.

As described in Line 13, **Algorithm 1**, the Q-function will be updated according to the Bellman equation by the Q-function at current time $Q(\mathbf{s}^{(k)}, \mathbf{x}^{(k)})$, the utility of the UAV $u^{(k)}$, and the value function of the next state denoted by $V\left(\mathbf{s}^{(k+1)}\right)$, which is the highest Q function over all possible policies for a given state and can be obtained by

$$V\left(\mathbf{s}^{(k+1)}\right) = \max_{\mathbf{x}} Q\left(\mathbf{s}^{(k+1)}, \mathbf{x}\right). \tag{3}$$

The learning rate denoted by $\alpha \in (0, 1]$ is used to represent the weight of the Q value at current time, and the discount factor representing the weight of the future utility is denoted by $\delta \in [0, 1]$. By iteratively updating the Q function, the optimal transmission policy can be learnt in the RL-based framework, which optimizes the video QoE, energy consumption, and utility.

V. Performance Analysis

We derive and analyze the convergence performance of the proposed RUMT scheme with regard to the video quality, energy consumption and utility. The superscript k is omitted without confusion. For simplicity, the PSNR is modeled with $\omega a^{\beta p} (\omega > 0, \beta < 0, a > 1)$, where $p \geq 0$ is the value of QP, and ω, a and β are fitting parameters [24]. The channel power gain in dB can be estimated according to the air-to-ground channel path loss model [22] given by

$$g = -\frac{|d|^{\kappa} \left(1 + \varepsilon \lambda \exp(-\mu(\theta - \lambda))\right)}{1 + \lambda \exp\left(-\mu(\theta - \lambda)\right)} (\text{dB}), \tag{4}$$

where λ and μ are constant parameters dependent on the environment, κ is the path loss exponent and ε is the additional path loss factor of the NLoS link. The end-to-end delay including the processing delay and transmission delay can be modelled with

$$\tau = \phi p^2 + \eta p + \nu + \frac{N_{\text{B}}}{b}, \tag{5}$$

where $\phi > 0$, $\eta < 0$ and $\nu > 0$ are fitting parameters [15] and N_{B} is the number of the compressed video data bits.

Theorem 1. *The proposed RUMT scheme can achieve an optimal performance consisting of the video quality and energy consumption by $\omega a^{\beta P_1}$ and $X_1 T_{\text{tx}}$, respectively, and an optimal utility by*

$$u = \frac{\left(g - \xi T_{\text{tx}} \sigma_n^2\right) X_1}{\sigma_n^2} + \omega a^{\beta P_1} - \varphi l \left(\phi P_1^2 + \eta P_1 + \nu + \frac{N_{\text{B}}}{b}\right), \tag{6}$$

if

$$-\sqrt{2\varphi l \phi \omega^{-1}} < \beta \ln a < \varphi l \eta \omega^{-1}, \tag{7}$$

$$g < \xi T_{\text{tx}} \sigma_n^2. \tag{8}$$

Proof: If (7) holds, we have

$$\frac{\partial^2 u}{\partial p^2} = \omega (\beta \ln a)^2 a^{\beta p} - 2\varphi l \phi \leq \omega (\beta \ln a)^2 - 2\varphi l \phi < 0,$$

$$\begin{aligned}\frac{\partial u}{\partial p} &= \omega\beta(\ln a)a^{\beta p} - 2\varphi l\phi p - \varphi l\eta \\ &\le \omega\beta\ln a - \varphi l\eta < 0, \end{aligned} \tag{9}$$

indicating that the utility decreases with p.

If (8) holds, we have

$$\frac{\partial u}{\partial x} = \frac{g - \xi T_{\text{tx}}\sigma_n^2}{\sigma_n^2} < 0, \tag{10}$$

indicating that the utility decreases with x. As $p \in [P_1, P_2]$ and $x \in [X_1, X_2]$, we have the optimal transmission policy given by $\mathbf{x}^* = [P_1, X_1] = \arg\max_{\mathbf{x}} u$.

This proposed RUMT scheme can achieve the optimal transmission policy $\mathbf{x}^* = [P_1, X_1]$ in the dynamic environment modeled as an MDP after a sufficient long time according to [23]. Therefore, we can derive (6). The optimal video quality and energy consumption can also be obtained. ■

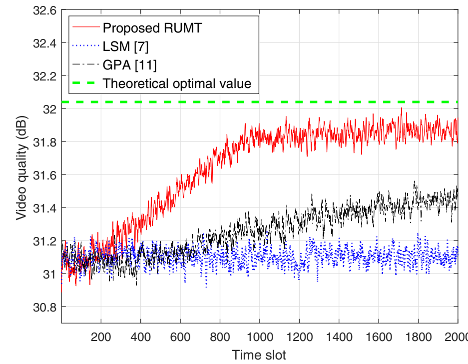
Remark 1: The UAV applies the RL-based media transmission algorithm to select the optimal transmission policy with the unknown specific video transmission model. If the UAV has the video quality with proper β as shown in (7), the utility of the UAV decreases with QP as shown in (9), so the UAV should choose the minimum QP to quantize the captured video precisely. If the UAV-CS link suffers from bad channel conditions with low channel power gain as shown in (8), the utility of the UAV is inversely correlated to the transmit power as shown in (10) so the UAV should transmit the compressed video with the minimum transmit power to maximize its utility.

VI. Simulation Results

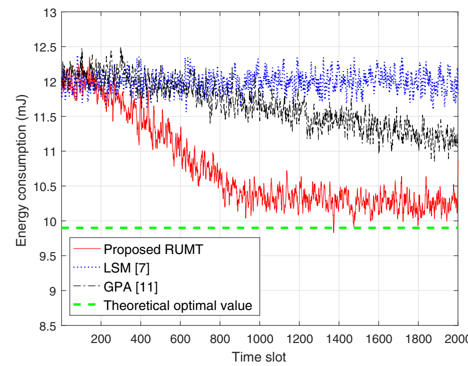
The performance of the proposed UAV-aided media transmission scheme based on RL has been evaluated through simulations, in which the UAV located at the coordinate of $(1, 1, 2)$ at the initial time compresses the captured video and transmits it to the CS located at $(0, 0, 0)$. The available channel bandwidth is uniformly distributed from 1 MHz to 3 MHz. The task priority is divided into 4 levels from 0 to 3 to represent the emergency level of an event. The path loss exponent κ and the additional path loss factor of the NLoS link ε are set to 3 and 20, respectively [22].

The QP is discretized into 6 levels from 0 to 51 with the step of 10, and the transmit power is discretized into 5 levels from 60 to 180 mW with the step of 30 mW. Unless specified otherwise, the learning rate is set to 0.7 and the discount factor is set to 0.8 in **Algorithm 1**. The video transmission time T_{tx} is set to 100 ms. As a benchmark, the LSM-based scheme determines the value of QP by $p^{(k)} = \ln(T_{\text{R}}/\vartheta)$, where T_{R} is the total number of retransmission bits and ϑ is the coding complexity, and a fixed QP with several specific values is used [7]. The GPA scheme [11] is evaluated in the simulations as another benchmark, which selects the optimal allocated power for the current state at each step.

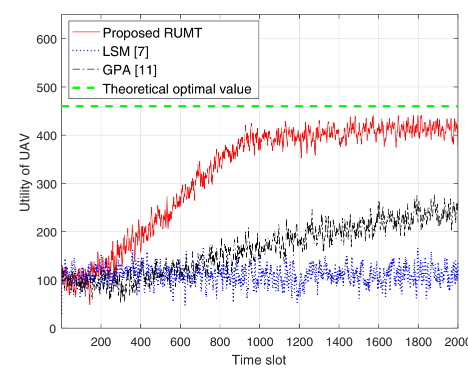
Simulation results in Fig. 2 show that the proposed RL-based media transmission scheme for UAV-aided networks is able to select the QP and transmit power to obtain the optimal utility via exploitation and exploration, which significantly outperforms the benchmark schemes in video quality, energy consumption and utility. The theoretical optimal value derived in Section V is also provided in the results.



(a) Video quality at the CS



(b) Energy consumption of the UAV



(c) Utility of the UAV

Fig. 2: Performance of the RUMT scheme with $\alpha = 0.7$ and $\delta = 0.8$ and two benchmarks.

As shown in Fig. 2 (a), the video quality of the proposed RUMT scheme is increased by about 20.2% over time from 31.1 dB at the beginning to 31.9 dB at convergence after 1000 time slots, while the LSM-based scheme almost keeps the video quality of 31.1 dB and the GPA scheme is limited to 31.2 dB. Thus the video quality of the proposed scheme is improved by 20% and 17.5% at the 1000-th time slot compared with the LSM-based scheme and the GPA scheme, respectively.

As is shown in Fig. 2 (b), the energy consumption of the proposed RUMT scheme is decreased by about 14.2%, from 12 mJ at the beginning to 10.3 mJ, after 1000 time slots at convergence, while the LSM-based scheme hardly changes and keeps the energy consumption of 12.1 mJ and the GPA scheme is decreased to 11.6 mJ. Thus the energy consumption can be decreased by 14.9% and 11.2% at the 1000-th time slot using the proposed scheme compared with the LSM-based scheme and the GPA scheme, respectively.

As is shown in Fig. 2 (c), the utility of the proposed RUMT scheme is significantly increased by about 281.9% from 105 at the beginning to 401 at convergence after 1000 time slots, while the LSM-based scheme approximately maintains the utility of 110 and the GPA scheme is increased up to 167. Thus the utility of the proposed scheme outperforms the LSM-based scheme and the GPA scheme by 264.5% and 140.1% at the 1000-th time slot, respectively.

Consequently, it can be validated from the simulation results that the proposed RUMT scheme significantly improves the video transmission performance and the video QoE compared with the existing advanced benchmarks. Furthermore, it can be noted from Fig. 2 that the proposed scheme is approaching the theoretical optimal value derived in Section V, which verifies the effectiveness of the proposed media transmission scheme using RL algorithms.

VII. Conclusion

In this paper, we have proposed an RL-based UAV-aided media transmission scheme to improve the transmission performance and the video QoE, such as the video quality, end-to-end delay, and energy consumption. Through the RL process, the UAV dynamically selects the QP to compress the video of the target area, and transmits the processed video using the selected transmit power to the CS via trail and error without knowing the specific video transmission model. It has been proved that the RUMT scheme can obtain the optimal performance in the dynamic UAV-aided networks, and the performance bound regarding the video quality, energy consumption and utility has been provided. The video transmission performance has been evaluated through simulations, showing that the proposed scheme is significantly improved compared with the state-of-the-art LSM-based scheme and GPA scheme. Furthermore, the proposed scheme can hopefully be applied in various areas on smart and efficient media transmission.

References

[1] Y. Zhou, N. Cheng, N. Lu, *et al.*, "Multi-UAV-aided networks: Aerial-ground cooperative vehicular networking architecture," *IEEE Veh. Technol. Mag.*, vol. 10, no. 4, pp. 36–44, Dec. 2015.

[2] J. Zhang, J. Weng, W. Luo, *et al.*, "REMT: A real-time end-to-end media data transmission mechanism in UAV-aided networks," *IEEE Network*, vol. 32, no. 5, pp. 118–123, Apr. 2018.

[3] N. Cheng, W. Xu, W. Shi, *et al.*, "Air-ground integrated mobile edge networks: Architecture, challenges, and opportunities," *IEEE Commun. Mag.*, vol. 56, no. 8, pp. 26–32, Aug. 2018.

[4] J.-R. Ohm, G. Sullivan, H. Schwarz, *et al.*, "Comparison of the coding efficiency of video coding standards-including high efficiency video coding (HEVC)," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 22, no. 12, pp. 1669–1684, Dec. 2012.

[5] Y.-F. Ou, Y. Xue, and Y. Wang, "Q-STAR: A perceptual video quality model considering impact of spatial, temporal, and amplitude resolutions," *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2473–2486, Jun. 2014.

[6] L. Xu, W. Lin, L. Ma, *et al.*, "Free-energy principle inspired video quality metric and its use in video coding," *IEEE Trans. Multimedia*, vol. 18, no. 4, pp. 590–601, Apr. 2016.

[7] M. Usman, N. Yang, M. Jan, *et al.*, "A joint framework for QoS and QoE for video transmission over wireless multimedia sensor networks," *IEEE Trans. Mobile Computing*, vol. 17, no. 4, pp. 746–759, Apr. 2018.

[8] Q. Wu, Y. Zeng, and R. Zhang, "Joint trajectory and communication design for multi-UAV enabled wireless networksy," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2109–2121, Mar. 2018.

[9] H. Mukhtar, A. Al-Dweik, M. Al-Mualla, *et al.*, "Low complexity power optimization algorithm for multimedia transmission over wireless networks," *IEEE J. Sel. Top. Signal Process.*, vol. 9, no. 1, pp. 113–124, Feb. 2015.

[10] L. Kaelbling, M. Littman, and A. Moore, "Reinforcement learning: A survey," *Journal of artificial intelligence research*, vol. 4, no. 11, pp. 237–285, May 2016.

[11] H. Cui, D. Liu, Y. Han, *et al.*, "Robust uncoded video transmission under practical channel estimation," in *IEEE Global Commun. Conf. (GLOBECOM)*, pp. 1–6, Washington, DC, Dec. 2016.

[12] M. Bhaskaranand and J. Gibson, "Global motion assisted low complexity video encoding for UAV applications," *IEEE J. Sel. Top. Signal Process.*, vol. 9, no. 1, pp. 139–150, Feb. 2015.

[13] J. Wu, c. Yuen, N.-M. Cheung, *et al.*, "Enabling adaptive high-frame-rate video streaming in mobile cloud gaming applications," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 25, no. 12, pp. 1988–2001, Dec. 2015.

[14] Z. Pan, Y. Zhang, and S. Kwong, "Efficient motion and disparity estimation optimization for low complexity multiview video coding," *IEEE Trans. Broadcasting*, vol. 61, no. 2, pp. 166–176, Jun. 2015.

[15] H. Hassan, M. Khan, S. Gilani, *et al.*, "H.264 encoder parameter optimization for encoded wireless multimedia transmissions," *IEEE Access*, vol. 6, pp. 22046–22053, Apr. 2018.

[16] B. Li, H. Li, L. Li, *et al.*, "λ domain rate control algorithm for high efficiency video coding," *IEEE Trans. Image Process.*, vol. 23, no. 9, pp. 3841–3854, Sept. 2014.

[17] Y. Eisenberg, C. Luna, T. Pappas, *et al.*, "Joint source coding and transmission power management for energy efficient wireless video communications," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 12, no. 6, pp. 411–424, Jun. 2002.

[18] H. Wang, G. Ding, F. Gao, *et al.*, "Power control in UAV-supported ultra dense networks: Communications, caching, and energy transfer," *IEEE Commun. Mag.*, vol. 56, no. 6, pp. 28–34, Jun. 2018.

[19] B. Cheng, J. Yang, S. Wang, *et al.*, "Adaptive video transmission control system based on reinforcement learning approach over heterogeneous networks," *IEEE Trans. Automation Science and Engineering*, vol. 12, no. 3, pp. 746–759, Jul. 2015.

[20] C. Wang, J. Wang, and X. Zhang, "Autonomous navigation of UAV in large-scale unknown complex environment with deep reinforcement learning," in *IEEE Global Conf. Signal and Information Processing (GlobalSIP)*, pp. 858–862, Montreal, QC, Canada, Nov. 2017.

[21] A. Bokani, S. Hoseini, M. Hassan, *et al.*, "Implementation and evaluation of adaptive video streaming based on Markov decision process," in *IEEE Int. Conf. Commun. (ICC)*, pp. 1–6, Kuala Lumpur, Malaysia, May 2016.

[22] F. Zeng, R. Zhang, and X. Cheng, "UAV-assisted data dissemination scheduling in VANETs," in *IEEE Int. Conf. Commun. (ICC)*, pp. 1–8, Kansas City, MO, May 2018.

[23] R. Sutton and A. Barto, *Reinforcement learning: An introduction*. MIT press, Cambridge, MA, 1998.

[24] S. Ma, W. Gao, and Y. Lu, "Rate-Distortion analysis for H.264/AVC video coding and its application to rate control," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 15, no. 12, pp. 1533–1544, Dec. 2005.