

基于强化学习的能量采集物联网视频编码技术

黄锦灏¹, 江东华¹, 丁钰真¹, 肖亮^{1*}, 范业仙², 陈建成³

(1. 厦门大学信息科学与技术学院, 福建 厦门 361005; 2. 宁德师范学院信息与机电工程学院, 福建 宁德 352100;
3. 厦门盈趣科技股份有限公司, 福建 厦门 361006)

摘要: 物联网设备的视频压缩码率和量化参数等编码方案决定着视频质量和物联网设备的使用寿命. 针对未知视频传输模型和能量采集模型的物联网设备, 提出一种基于强化学习的视频编码方案, 能够动态优化视频编码码率和量化参数. 该技术根据物联网设备的无线信道带宽、电池水平和采集的能量, 结合反馈的视频质量和时延等信息, 采用强化学习算法优化选择视频编码码率和量化参数. 在动态的网络环境下, 物联网设备不需预知视频传输模型就可以综合优化视频质量和设备能量损耗. 仿真结果表明, 该方案可以提高视频质量, 降低设备能量损耗和时延, 改善物联网设备效益.

关键词: 物联网; 视频编码; 强化学习; 能量采集

中图分类号: TN 911.21

文献标志码: A

文章编号: 0438-0479(2019)05-0889-05

物联网技术被广泛应用于医疗保健^[1]、智能交通^[2]、环境检测^[3]等领域, 其视频压缩编码和传输方案决定了物联网设备的能量损耗、视频传输质量和用户的网络体验^[4]. 随着第五代移动通信技术^[5]和边缘计算^[6]的发展, 物联网用户广泛关注传输的安全性^[7]、隐私性^[8]、用户体验度和物联网设备的电池寿命^[9]. 同时, H.264 等视频编码标准虽可改善视频的压缩比和对信道的适应性^[10], 但却增加了传输能耗^[11]和要求更短的视频传输时延^[12]. 因此, 如何优化物联网设备的视频编码方案, 综合提高视频质量、降低设备能耗和时延, 成为物联网技术发展的关键问题之一.

能量采集技术通过采集装置将射频信号等环境能量转化为电能, 提高了网络设备的电池使用寿命^[13]. 例如, 物联网设备根据当前电量水平和能量采集效率优化中继选择以降低成本^[14]或者在任务调度中优化采集能量的分配可以执行更多的任务^[15]. 目前, 物联网设备基于已知的视频传输模型和无线信道模型等信息来优化视频压缩编码的量化参数^[16]和编码码率^[17], 从而提高视频传输质量. 然而, 能量采集物联网设备在动态的网络环境中难以精确预估上述信息并据此来优化视频质量. 物联网设备应用强化学习技术可以

在动态博弈中优化其策略, 提高设备的传输质量等效益. 例如, 物联网设备采用强化学习为用户不同的需求分配缓存资源, 不需要预知网络模型即可提高用户的体验质量^[18]; 物联网医疗设备采用强化学习选择感知数据的上传策略, 从而保证用户传输模型的隐私^[1].

因此, 本文中提出了一种基于强化学习的能量采集物联网设备视频编码参数选择方案. 该技术选择视频压缩的编码码率和量化参数, 将其卸载至边缘设备进行数据处理, 旨在提高视频质量, 降低计算和传输的能量损耗和时延, 并应用能量采集技术对物联网设备补给能量来延长电池寿命. 该过程可建模为马尔可夫决策过程, 物联网设备根据边缘设备反馈的视频质量和时延, 以及当前的无线信道带宽、设备电能和采集的能量等信息, 利用强化学习算法动态优化视频编码码率和量化参数. 该方案不需要预知视频传输模型和能量采集模型, 通过试错来改善视频质量、设备能耗和传输时延等视频传输性能.

1 系统模型

考虑一个能量采集物联网视频传输系统由物联

收稿日期: 2019-01-24 录用日期: 2019-07-10

基金项目: 国家自然科学基金(61671396); 福建省自然科学基金(2019J01843); 东南大学移动通信国家重点实验室开放基金(2018D08).

* 通信作者: lxiao@xmu.edu.cn

引文格式: 黄锦灏, 江东华, 丁钰真, 等. 基于强化学习的能量采集物联网视频编码技术[J]. 厦门大学学报(自然科学版), 2019, 58(5): 889-893.

Citation: HUANG J H, JIANG D H, DING Y Z, et al. Video coding for energy harvested IoT devices based on reinforcement learning[J]. J Xiamen Univ Nat Sci, 2019, 58(5): 889-893. (in Chinese)



<http://jxmu.xmu.edu.cn>

网设备和边缘设备组成. 物联网设备首先选择合适的视频编码码率和量化参数对视频进行编码, 然后将编码后视频传输给边缘设备, 最后获得边缘设备反馈的视频质量和时延. 在这个过程中应用能量采集技术来对物联网设备提供能量, 提高物联网设备完成视频传输任务的概率.

1.1 网络模型

如图 1 所示, 物联网设备在 k 时刻与边缘设备之间的信道特征为 $g^{(k)}$, 自身电能为 $b^{(k)}$, 可用的传输信道带宽为 $B^{(k)}$. 为了增加电能, 设物联网设备利用射频天线或者太阳能电池从环境中采集的能量为 $\rho^{(k)}$. 根据 H. 264 视频编码标准, 物联网设备选择视频编码码率 $x_1^{(k)}$ 和量化参数 $x_2^{(k)}$ 对将要发送的视频进行编码处理, 并将处理后的视频传输给边缘设备进行解压和其他进一步的处理. 边缘设备在接收到处理压缩的视频后将视频质量和时延反馈给物联网设备.

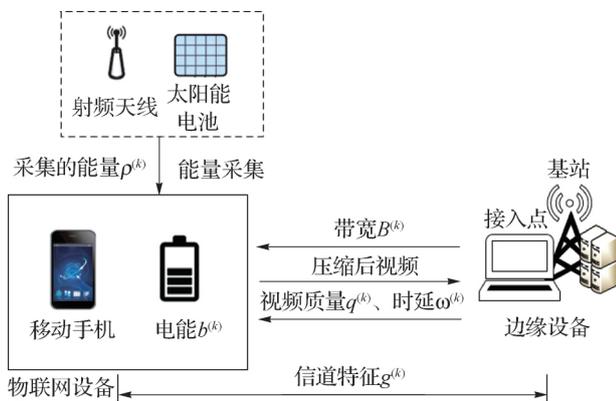


图 1 能量采集物联网视频传输模型
Fig. 1 Video transmission model for IoT devices with energy harvesting

1.2 视频传输模型

视频编码传输过程主要考虑视频质量、能量损耗和时延. 视频质量 $q^{(k)}$ 包含了视频的压缩质量 $\eta^{(k)}$ 和传输质量 $\nu^{(k)}$, 即 $q^{(k)} = \eta^{(k)} + \nu^{(k)}$. 压缩质量 $\eta^{(k)}$ 通过编码后视频的峰值信噪比 (peak signal-to-noise ratio, PSNR) 来衡量. PSNR 是一种常用的客观视频质量评价指标, 值越大表示视频质量越高. 传输质量 $\nu^{(k)}$ 使用信噪比 (signal-to-noise ratio, SNR) 来进行衡量. 视频传输过程中的能量损耗 $E^{(k)}$ 会影响物联网设备的电能和任务完成情况. 为了避免电量不足导致的任务失败, 物联网设备利用能量采集技术进行自产能.

视频传输的时延 $\omega^{(k)}$ 包括排队时延、处理时延、传输时延和传播时延 4 个部分, 根据文献[19], 视频传

输过程所需带宽不会超过系统总有效带宽, 并且实验证明排队时延不会影响视频传输质量, 而传播时延和其他部分相比较小, 所以在本方案中不考虑将排队时延和传播时延作为优化目标, 只关注处理时延 (即视频压缩编码时间) 和传输时延这两个部分.

2 基于强化学习的能量采集物联网设备视频编码选择方案

由于能量采集物联网设备的视频编码传输过程可以建模为马尔可夫决策过程, 为了得到视频编码的最优参数选择, 物联网设备利用强化学习优化视频编码传输过程, 动态选择视频编码码率 $x_1^{(k)}$ 和量化参数 $x_2^{(k)}$, 从而提高视频质量并降低能量损耗和时延.

设物联网设备在接收到边缘设备反馈的上一时刻 $q^{(k-1)}$ 和 $\omega^{(k-1)}$ 后, 观测 $g^{(k)}$ 、 $b^{(k)}$ 、 $\rho^{(k)}$ 和可用的传输信道带宽 $B^{(k)}$, 将视频编码传输过程建模为马尔可夫决策过程则有系统状态空间 $s^{(k)} = [\rho^{(k)}, b^{(k)}, B^{(k)}, g^{(k)}, q^{(k-1)}, \omega^{(k-1)}]$ 和动作空间 $x^{(k)} = [x_1^{(k)}, x_2^{(k)}]$. 使用 ϵ 贪婪策略动态选择 $x^{(k)}$, 即用 $1-\epsilon$ 的大概率选择满足最大化 Q 函数的 $x^{(k)}$, 并保留极小概率 ϵ 进行探索, 随机选择其他动作 $x^{(k)}$ 视频编码策略, 其中 $0 < \epsilon < 1$.

根据视频编码传输过程中的 $q^{(k)}$ 、 $E^{(k)}$ 、 $\omega^{(k)}$ 和任务失败的额外损耗评估物联网设备的效益 $u^{(k)}$, 定义单位能量损耗 σ 、单位时延 τ 、额外损耗系数 γ , 则有

$$u^{(k)} = q^{(k)} - \sigma E^{(k)} - \tau \omega^{(k)} - \gamma \phi(b^{(k)} - E^{(k)} + \rho^{(k)} < 0). \tag{1}$$

其中, ϕ 为指示函数, 表示当 $b^{(k)} - E^{(k)} + \rho^{(k)} < 0$ 时物联网设备传输任务失败需要付出的额外损耗. 另外, Q 函数 $Q(s^{(k)}, x^{(k)})$ 根据当前动作状态对 $(s^{(k)}, x^{(k)})$ 、当前时刻的效益 $u^{(k)}$ 和下一时刻的值函数 $V(s^{(k+1)})$ 使用学习速率 α 和折扣因子 δ 进行更新, 其中学习速率 $\alpha \in (0, 1]$, 折扣因子 $\delta \in (0, 1]$.

基于强化学习的能量采集物联网设备的视频编码参数选择技术的算法步骤如下:

- 1) 初始化强化学习的参数: 学习速率 α , 折扣因子 δ , 初始状态 $s^{(0)}$, Q 函数 $Q(s^{(0)}, x^{(0)}) = 0$ 和值函数 $V(s^{(1)}) = 0$.
- 2) 根据强化学习对 Q, V 进行迭代运算.
 - a) 接收边缘设备反馈的上一时刻视频质量 $q^{(k-1)}$ 和时延 $\omega^{(k-1)}$;
 - b) 观测自身电能 $b^{(k)}$ 、传输信道带宽 $B^{(k)}$ 、信道特征 $g^{(k)}$ 和采集的能量 $\rho^{(k)}$;
 - c) 构建当前时刻的系统状态 $s^{(k)} = [\rho^{(k)}, b^{(k)},$

$B^{(k)}, g^{(k)}, q^{(k-1)}, \omega^{(k-1)}];$

d) 根据 ϵ 贪婪策略选取视频编码策略 $x^{(k)} = [x_1^{(k)}, x_2^{(k)}];$

e) 利用视频编码码率 $x_1^{(k)}$ 和量化参数 $x_2^{(k)}$ 对视频进行编码处理;

f) 传输编码后的视频到边缘设备;

g) 观测能量损耗 $E^{(k)}$ 和下一时刻电能 $b^{(k+1)}$;

h) 根据公式(5)评估物联网设备的效益 $u^{(k)}$;

i) 根据 $Q(s^{(k)}, x^{(k)}) \leftarrow (1 - \alpha)Q(s^{(k)}, x^{(k)}) + \alpha(u^{(k)} + \delta V(s^{(k+1)}))$ 更新 Q 函数 $Q(s^{(k)}, x^{(k)});$

j) 根据 $V(s^{(k)}) = \max Q(s^{(k)}, x^{(k)})$ 更新值函数 $V(s^{(k)});$

3) 当 $k=K$ 时停止迭代运算。

3 仿真分析

本章利用 MATLAB 软件,对提出的基于强化学习的能量采集物联网设备的视频编码参数选择方案进行仿真分析.物联网设备与边缘设备在无线空间环境中进行视频传输,主要存在以地面反射波为主导的多径效应,在此应用两径模型^[20]对传输信道进行建模.两径模型的接收信号包含经自由空间到达接收端的直射分量和地面反射分量两个部分.具体来说,信道特征 $g^{(k)}$ 与电磁波波长 λ 、物联网设备高度 h_M 、边缘设备的高度 h_D 、物联网设备和边缘设备之间的距离 $d^{(k)}$ 以及信道阴影衰落因子 φ 有关,其中信道阴影衰落因子 φ 服从高斯分布,其均值为 μ_0 ,方差为 σ_0^2 ^[21].信道特征 $g^{(k)}$ 可用如下公式进行评估:

$$g^{(k)} = \begin{cases} \frac{\lambda^2}{10^{0.1\varphi} (4\pi d^{(k)})^2}, & d^{(k)} \leq \frac{4\pi h_M h_D}{\lambda} \\ \frac{h_M^2 h_D^2}{10^{0.1\varphi} (d^{(k)})^4}, & d^{(k)} > \frac{4\pi h_M h_D}{\lambda} \end{cases} \quad (2)$$

其中,电磁波波长 λ 设为 5.9×10^9 m,物联网设备高度 h_M 设为 1.2 m,边缘设备的高度 h_D 设为 1.2 m.根据参考文献[22],编码后视频的 PSNR 可表示为 $\eta^{(k)} = ae^{-bx^{(k)}}$,其中 a 和 b 均为系统模型参数.根据参考文献[23]视频传输的 SNR 可表示为 $\nu^{(k)} = (c_0 + c_1(d^{(k)})^{c_2}) x_1^{(k)} g^{(k)} / P_n$,其中: P_n 表示服从高斯分布的噪声功率,其均值为 μ_1 ,方差为 σ_1^2 ; c_0 为传输损耗系数, c_1 为传输放大器系数, c_2 为路径损耗指数.故

$$q^{(k)} = ae^{-bx^{(k)}} + (c_0 + (c_1 d^{(k)})^{c_2}) x_1^{(k)} g^{(k)} P_n^{-1}. \quad (3)$$

根据文献[22-24],物联网设备可选的视频编码码率 $x_1^{(k)}$ 范围为 8~16 Mbit/s,步长为 2 Mbit/s,系统模型参数 a 为 35, b 为 -0.017,传输损耗系数 c_0 为 0.5,传

输放大器系数 c_1 为 1.3×10^{-8} ,路径损耗指数 c_2 为 4.能量损耗可用如下公式表示:

$$E^{(k)} = [c_0 + (c_1 d^{(k)})^{c_2}] x_1^{(k)} t. \quad (4)$$

其中,视频传输时间 t 为 300 ms,根据文献[25],视频传输所需时延为:

$$\omega^{(k)} = [\omega_1 (x_2^{(k)})^2 + \omega_2 x_2^{(k)} + \omega_3] + (B^{(k)})^{-1} N_B.$$

其中,系统模型参数 ω_1 为 0.01, ω_2 为 -1.37, ω_3 为 8.3.视频传输的比特数 N_B 为 128 bit.在本文算法中,设置 α 为 0.7, δ 为 0.8.式(1)中的单位能量损耗系数 σ 为 1,单位时延系数 τ 为 8,额外损耗系数 γ 为 200.为了评估本文算法的性能,对比了文献[26]提出的基于拉格朗日乘子(lagrange multiplier, LM)、忽略无线信道衰落的强化学习(Q-learning)和考虑无线信道衰落的强化学习(Q-learning-shadow)3种方法的视频质量、物联网设备的效益、能量损耗和时延.

从图2可看出,无线信道衰落对本文强化学习方法的影响不大.忽略无线信道衰落时,相较于基于 LM 的方案,本文中提出的基于强化学习的能量采集物联网设备的视频编码方案能够使传输的视频质量从 33.8 dB 上升到 34.9 dB,并在 3 s 时达到收敛,相较于基于 LM 的方案,本文中提出的方案能使传输的视频质量提高 29.2%(图2(a)).使物联网设备的效益从 -1 000 上升到 300,并在 3 s 时达到收敛,与基于 LM 的方案相比,能使物联网设备的效益提高 13 倍(图2(b));能量损耗从 1.65 J 下降到 1.40 J,并在 3 s 时达到收敛.相较于基于 LM 的方案,本文中提出的方案能使物联网设备的能量损耗下降 13.3%;能使物联网设备的时延从 1.60 s 下降到 1.20 s,并在 3 s 时达到收敛,与基于 LM 的方案相比,能使视频传输所需时延下降 25%(图2(d)).

综上所述,本文中所提基于强化学习的能量采集物联网设备的视频编码方案性能明显优于基于 LM 的方案,即能够更有效地提高视频质量和物联网设备的效益,并且降低能量损耗和时延.

4 结论

本文中提出了一种基于强化学习的能量采集物联网设备的视频编码技术,物联网设备通过采集外界能量并转化为自身电量来抵消部分能量损耗,并利用强化学习技术进行视频编码参数选择,根据传输信道带宽、电能、上一时刻视频质量等系统状态动态选择视频编码码率和量化参数,提高视频传输性能.仿真结果表明,本文中提出的方案比基于 LM 的方案能更

<http://jxmu.xmu.edu.cn>

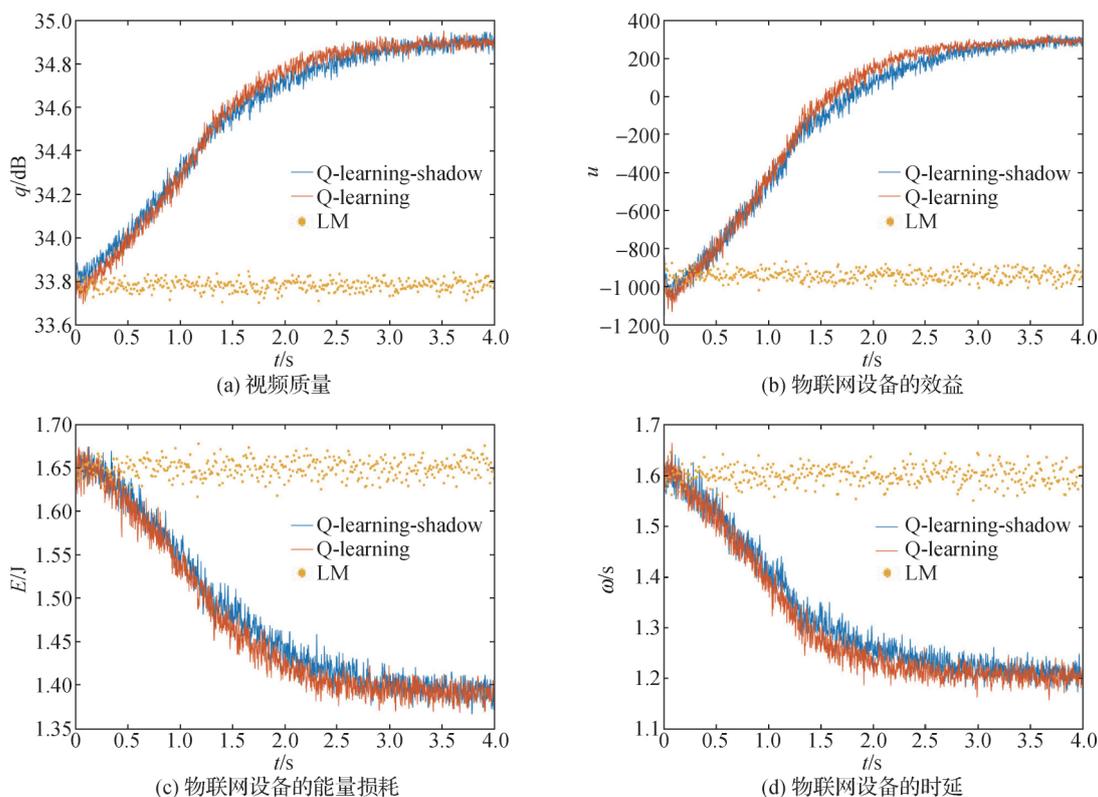


图 2 能量采集物联网设备视频编码参数选择结果

Fig. 2 The results of video coding parameter selection scheme for energy harvested IoT devices

有效地提高视频质量和物联网设备的效益,并且降低物联网设备的能量损耗和时延。

参考文献:

[1] MIN M, WAN X, XIAO L, et al. Learning-based privacy-aware offloading for healthcare IoT with energy harvesting [J]. IEEE Internet of Things Journal, 2019, 6(3): 4307-4316.

[2] CENTENARO M, VANGELISTA L, ZANELLA A, et al. Long-range communications in unlicensed bands: the rising stars in the IoT and smart city scenarios[J]. IEEE Wireless Communications, 2015, 23(5): 60-67.

[3] MOIS G, FOLEA S, SANISLAV T. Analysis of three IoT-based wireless sensors for environmental monitoring [J]. IEEE Transactions on Instrumentation & Measurement, 2017, 66(8): 2056-2064.

[4] 魏淑芝, 朱琦. 基于网络选择的视频通信带宽博弈算法 [J]. 通信学报, 2015, 36(2): 212-220.

[5] 王胡成, 徐晖, 程志密, 等. 5G 网络技术研究现状和发展趋势 [J]. 电信科学, 2015, 31(9): 149-155.

[6] XIAO L, WAN X, DAI C, et al. Security in mobile edge caching with reinforcement learning [J]. IEEE Wireless Communications, 2018, 25(3): 116-122.

[7] XIAO L, ZHUANG W, ZHOU S, et al. Learning-based

VANET communication and security techniques[M]. Cham: Springer, 2019: 13-45.

[8] XIAO L, WAN X, LU X, et al. IoT security techniques based on machine learning [J]. IEEE Signal Processing Magazine, 2018, 35(5): 41-49.

[9] KAUR N, SOOD S K. An energy-efficient architecture for the internet of things (IoT) [J]. IEEE Systems Journal, 2017, 11(2): 796-805.

[10] 朱秀昌. 视频编码新标准——H. 264 [J]. 电信科学, 2002, 18(12): 26-29.

[11] XIAO L, CHEN T, XIE C, et al. Mobile crowdsensing games in vehicular networks [J]. IEEE Transactions on Vehicular Technology, 2018, 67(2): 1535-1545.

[12] POULARAKIS K, IOSIFIDIS G, ARGYRIOU A, et al. Video delivery over heterogeneous cellular networks: optimizing cost and performance [C] // IEEE International Conference on Computer Communications. Piscataway: IEEE, 2014: 1078-1086.

[13] CHEN G, ZHAN Y, CHEN Y, et al. Reinforcement learning based power control for in-body sensors in WBANs against jamming [J]. IEEE Access, 2018, 6: 37403-37412.

[14] KAWABATA H, ISHIBASHI K, VUPPALA S, et al. Robust relay selection for large-scale energy-harvesting

<http://jxmu.xmu.edu.cn>

- IoT networks [J]. IEEE Internet of Things Journal, 2017, 4(2): 384-392.
- [15] CARUSO A, CHESSA S, ESCOLAR S, et al. A dynamic programming algorithm for high-level task scheduling in energy harvesting IoT [J]. IEEE Internet of Things Journal, 2018, 5(3): 2234-2248.
- [16] SAHA A, LEE Y, HWANGY, et al. Context-aware block-based motion estimation algorithm for multimedia internet of things (IoT) platform [J]. Personal and Ubiquitous Computing, 2017, 22(1): 163-172.
- [17] XU J, ANDREPOULOS Y, XIAO Y, et al. Non-stationary resource allocation policies for delay-constrained video streaming: application to video over internet-of-things-enabled networks [J]. IEEE Journal on Selected Areas in Communications, 2014, 32(4): 782-794.
- [18] HE X, WANG K, HUANG H, et al. Green resource allocation based on deep reinforcement learning in content-centric IoT [J]. IEEE Transactions on Emerging Topics in Computing, 2018: 2805718. doi: 10.1109/TETC.2018.2805718.
- [19] LIU Y, NIU D, LI B. Delay-optimized video traffic routing in software-defined interdatacenter networks [J]. IEEE Transactions on Multimedia, 2016, 18(5): 865-878.
- [20] SOMMER C, DRESSLER F. Using the right two-ray model? A measurement-based evaluation of PHY models in VANETs [C] // ACM Conference on Mobile Ad Hoc Networking and Computing, New York: ACM, 2011: 1-3.
- [21] SAMIMI M K, RAPPAPORT T S, MACCARTNEY G R. Probabilistic omnidirectional path loss models for millimeter-wave outdoor communications [J]. IEEE Wireless Communications Letters, 2015, 4(4): 357-360.
- [22] ZHOU Y, SUN Y, YIN X, et al. Effective frame level rate control for H. 264/AVC video coding [C] // IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE, 2008: 1-5.
- [23] KIM S, LEE H, JEON D, et al. Reduction in encoding redundancy for overlapped FOVs over wireless visual sensor networks [J]. Digital Signal Processing, 2017, 50: 135-149.
- [24] CZUNI L, CSASZAR G, LICZAR A. Estimating the optimal quantization parameter in H. 264 [C] // International Conference on Pattern Recognition (ICPR). Piscataway: IEEE, 2006: 330-333.
- [25] HASSAN H, KHAN M, GILANI S, et al. H. 264 encoder parameter optimization for encoded wireless multimedia transmissions [J]. IEEE Access, 2018, 6: 22046-22053.
- [26] KUO C, SHIH Y, YANG S. Rate control via adjustment of Lagrange multiplier for video coding [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2016, 26(11): 2069-2078.

Reinforcement-learning-based video coding for energy harvested IoT devices

HUANG Jinhao¹, JIANG Donghua¹, DING Yuzhen¹,
XIAO Liang^{1*}, FAN Yexian², CHEN Jiancheng³

(1. School of Informatics, Xiamen University, Xiamen 361005, China; 2. College of Information and Mechanical & Electrical Engineering, Ningde Normal University, Ningde 352100, China;
3. Xiamen Intretech Technology Co., Ltd, Xiamen 361006, China)

Abstract: Video coding technology can improve the video quality and save the energy consumption of Internet of Things (IoT) devices, in which the optimal video coding policy depends on the coding parameter selection. In this paper, we propose a reinforcement-learning-based video coding scheme for IoT devices with energy harvesting without knowing the video transmission model. The IoT device selects the encoding bit rate and quantization parameter according to the measured wireless channel bandwidth, battery level of IoT device, harvested energy, transmission distance and the previous video quality and delay. This scheme can achieve a trade-off between the video quality and energy consumption via trials and errors in the dynamic IoT network. Simulation results show that the proposed scheme can improve the video quality and the utility of IoT devices as well as reduce the energy consumption and delay.

Keywords: IoT; video coding; reinforcement learning; energy harvesting

<http://jxmu.xmu.edu.cn>