

基于反馈的自适应参考帧选择的率失真优化分析

彭强 杨天武 张庆明

(西南交通大学信息科学与技术学院 成都 610031)

摘要: 为了提高数字视频实时传输的容差能力, 该文提出了一种基于反馈的自适应参考帧选择方法。在传统参考帧选择优点的基础上, 设计了一种新的通过率失真技术优化选择参考帧的帧间帧预测结构, 分析了选择不同参考帧编码时的编码码率以及它对编码失真的影响。引入了信道代价均值的概念, 探讨了信道误差的扩散过程及其对信道失真的影响。在联合信源和信道的率失真优化框架内, 选择使得率失真代价最小的候选参考帧作为当前帧编码的参考帧, 有效地改善了视频传输质量, 并减少了编/解码器所需缓存的参考帧。仿真实验显示自适应参考帧选择的 PSNR 性能在不同码率和丢包率下均优于经典的 ACK 和 NACK 参考帧选择方法。

关键词: 视频编码; 差错恢复; 参考帧选择; 率失真优化

中图分类号: TN919.8

文献标识码: A

文章编号: 1009-5896(2009)06-1455-05

Rate-Distortion Analysis for Feedback-Based Adaptive Reference Picture Selection

Peng Qiang Yang Tian-wu Zhang Qing-ming

(School of Information Science & Technology, Southwest Jiaotong University, Chengdu 610031, China)

Abstract: A feedback-based Adaptive Reference Picture Selection (ARPS) is proposed for improving error resilience in real-time video transmission. The novel predictive structure which the reference picture selection can be optimized with rate-distortion method is designed by taking advantage of classic reference picture selection. The bit-rate of coding with difference reference pictures is analyzed, and its impact on source distortion is discussed. By first introducing the concept of average channel cost, the propagation of channel error and its impact on channel distortion is studied. Within joint source-channel rate-distortion optimization framework, the video picture that leads to the minimum rate-distortion cost is selected as the current reference picture, which can efficiently enhance the video transmission quality and decrease the number of reconstruction frame buffers of encoder/decoder. Simulation results show that the proposed ARPS algorithm outperforms the classical ACK and NACK RPS method in different bit-rates and lost-rates in PSNR.

Key words: Video coding; Error resilience; Reference Picture Selection(RPS); Rate-distortion optimization

1 引言

视频编码通常采用运动补偿预测技术来提高编码效率, 但是当视频流在 Internet 和无线移动等不可靠网络中传输时, 视频传输差错将导致解码出错并使差错在时域和空域上扩散。在反馈信道存在的情况下, 阻止视频差错扩散的典型方法是参考帧选择(Reference Picture Selection, RPS), 它是通过改变帧间预测来降低视频差错扩散的一种有效方法^[1]。编码器根据解码器反馈的视频帧接收状态, 通过选择解码器正确收到的视频帧作为参考进行编码从而阻止视频差错扩散。文献[2]提出了两种参考帧选择模式, 即 ACK (acknowledgment)模式和 NACK(noacknowledgment)模式。文献[3,4]分析了存在多条传输路径时的参考帧选择方法。

RPS 的 ACK 模式适合丢包严重的通信网络, 而 NACK

模式适合丢包较少的通信网络。当网络丢包情况变化时, 可采用两种模式的动态切换技术。虽然动态切换技术具有对网络丢包变化的较好适应能力, 但是因为没有结合率失真优化技术, 所以仍难以实现信源和信道编码的最佳平衡。同时编/解码器都需要缓存大量参考帧, 不适合硬件实现。率失真优化是平衡信源和信道编码的有效方法^[5]。本文在率失真优化框架内扩展了 RPS 的编码思想, 提出了适合不同信道特征的自适应参考帧选择(Adaptive Reference Picture Selection, ARPS), 通过参考帧的优化选择使得视频端到端的传输失真最小, 同时明显减少了编/解码器所需缓存的参考帧。

2 ARPS 的帧间预测方式

经典 RPS 的 ACK 模式和 NACK 模式如图 1 所示。在 ACK 模式中, 编码器始终使用已被解码器确认的正确视频帧作为当前帧的参考, 如果当前帧被正确接收, 视频差错扩

2008-06-02 收到, 2008-10-06 改回

国家自然科学基金(60672099)和教育部留学回国人员科研启动基金资助课题

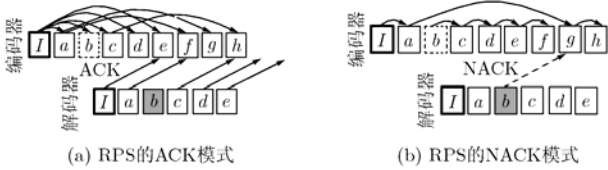


图1 经典参考帧选择(假定阴影帧出错)

散就被阻止。但是在视频传输中,往返时延一般是多个编码帧间隔,因此ACK模式的帧间预测效率较低,仅适合丢包严重的情况。在NACK模式中,编码器一般情况下进行正常编码,只有在收到解码器反馈的NACK信号时,编码器才选择比确认的丢失视频帧更早的正确帧作为参考帧进行编码。NACK模式提高了帧间预测效率,但是差错至少在往返时延间隔内进行扩散,因此NACK模式只适合丢包较少的情况。同时RPS的两种模式中的编/解码器都至少需要缓存往返时延内的所有视频帧。

本文提出的ARPS方法吸收了RPS两种模式的优点,根据信道误码情况自适应地选择不同的参考帧,其编码预测方式如图2所示。ARPS将所有视频帧分为两类,即普通帧和核心帧。图2中所有小写字母表示的视频帧均为普通帧,普通帧只参考前一帧进行编码,并在后一帧编码完成后自动释放。核心帧是编码器指定的长期帧(Long Term Picture, LTP),在图2中包括用I表示的帧内编码帧以及大写字母表示的帧间编码帧都是核心帧。

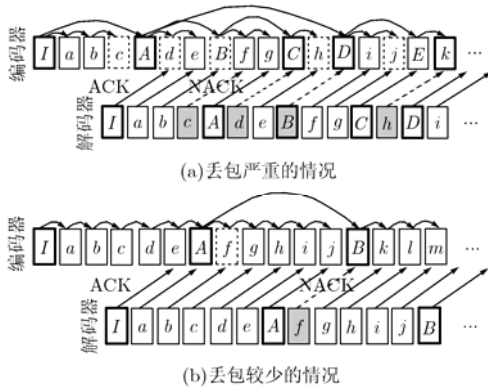


图2 自适应参考帧选择(假定阴影帧出错)

与经典的RPS不同,ARPS的解码器同时反馈ACK和NACK信息,但编码器不通过反馈信息直接控制参考帧选择,而是利用反馈信息估算出参考不同候选参考帧时的率失真代价,选择率失真代价最小的候选参考帧作为当前帧的参考帧。ARPS采用的候选参考帧包括前一帧、前一核心帧和前一正确核心帧。如果当前帧选择核心帧作为参考帧,则当前帧就被设置为新的核心帧。当编码器得知前一核心帧包含差错时,编码器则直接选择前一个正确核心帧作为当前帧的参考帧,例如在图2(a)中帧D编码时,收到了帧B的NACK

信息。因为前一核心帧C通过帧B预测编码,所以包含差错,因此当前帧直接参考前一正确核心帧A编码。当丢包率很低时,可能会连续编码多个普通帧,使得当前帧和前一正确核心帧的间隔逐渐变远,在一旦发生差错时,编码效率就会显著降低。为了提高编码效率,在当前帧前面5帧为普通帧时,就直接指定当前帧为核心帧,而当前帧仍可参考前一帧编码,例如图2(b)中的帧A。

ARPS编/解码器仅需要保存前一帧、前一正确核心帧,以及这两帧间的核心帧共3帧信息。对于解码器,如果一个视频帧完全丢失,则解码器不能判断它是否是核心帧,因此假定该帧是核心帧,需要在差错掩盖后缓存该帧。

3 自适应参考帧选择

3.1 编码代价

对于帧间编码,编码距离定义为当前帧与参考帧间的编码帧数。视频序列在没有镜头切换时,一般说来编码距离越长,当前帧与参考帧间的相关性就越低,运动补偿预测的效率就越差,即压缩质量不变时需要更多编码比特。实验表明编码数据率基本上与编码距离的对数保持线性关系,即

$$R(n, r) = (1 + k_Q \log_2(n - r))R(n, n - 1) \quad (1)$$

其中 $R(n, r)$ 表示当前帧第 n 帧参考第 r 帧编码时的编码比特率, k_Q 为常数,与量化步长相关,经验取值为 $0.18 + 0.01 \times QP_{H.264}$ 。为了证实上面关系,在JM9.8测试模型中测试了多个典型的视频序列,部分结果如图3所示。实验结果显示对于所有视频序列,在编码距离小于16帧时,线性关系都保持很好,当编码距离大于32帧时,个别序列的线性关系略有改变,但是大多数序列仍然保持线性关系,因此式(1)基本成立。

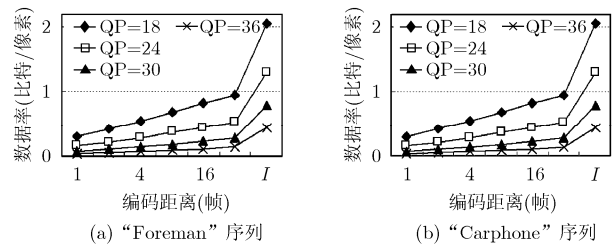


图3 编码码率与编码距离的关系

假定当前第 n 帧参考第 r 帧编码的编码比特为 $R(n, r)$,参考第 r' 帧编码的编码比特为 $R(n, r')$ 。则参考第 r 帧编码相对参考第 r' 帧编码增加的冗余编码比特为 $R(n, r) - R(n, r')$,它导致的编码失真增加 $\Delta d_Q(r, r')$ 为

$$\Delta d_Q(r, r') = \phi \times (R(n, r) - R(n, r')) \quad (2)$$

其中 ϕ 是冗余编码比特的拉格朗日系数,即每一冗余编码比特产生的信道失真。

3.2 端到端的传输失真

ARPS的关键是端到端的传输失真估计。假定当前第 n

帧像素 i 的原始值为 f_n^i 。它在编码器的重建值为 \hat{f}_n^i 。由于量化引起编码器重建值与原始值不同, 定义第 n 帧信道失真为

$$d_S(n) = \sum_{i \in \{\text{frame } n\}} [f_n^i - \hat{f}_n^i]^2 \quad (3)$$

像素 i 经不可靠网络传输后, 假定在解码器重建为 \tilde{f}_n^i 。因为视频差错的存在, 导致 \tilde{f}_n^i 与 \hat{f}_n^i 不完全相同。对于编码器而言, 因为无法正确确定 \tilde{f}_n^i , 所以将 \tilde{f}_n^i 看作为随机变量。定义像素 i 的传输失真 $d^i(n)$ 和信道失真 $d_C^i(n)$ 分别为

$$d(n) = \sum_{i \in \{\text{frame } n\}} E\{[f_n^i - \tilde{f}_n^i]^2\} \quad (4)$$

$$d_C(n) = \sum_{i \in \{\text{frame } n\}} E\{[\tilde{f}_n^i - \hat{f}_n^i]^2\} \quad (5)$$

文献[1]及本文的仿真实验结果都表明传输失真约等于信道失真与信源失真之和, 即 $d^i(n) \approx d_S^i(n) + d_C^i(n)$ 。当像素 i 未正确接收时, 假定采用第 $n-1$ 帧像素 k 进行差错掩盖, 定义第 n 帧的掩盖失真为

$$d_Y(n) = \sum_{i \in \{\text{frame } n\}} E\{[\tilde{f}_{n-1}^k - \tilde{f}_n^i]^2\} \quad (6)$$

如果假设丢包率为 p , 帧内宏块更新率为 α , 根据文献[1, 6]的结论, 如果第 n 帧选择第 l 帧为参考帧, 第 n 帧的信道失真可在编码前递归估计为

$$d_C(n) = (1-p)\alpha\beta d_C(l) + \beta d_C(n-1) + p\bar{d}_Y \quad (7)$$

其中 \bar{d}_Y 为掩盖失真均值, 因为在第 n 帧编码前, 无法通过差错掩盖仿真计算出掩盖失真, 所以采用 \bar{d}_Y 估计。 β 是信道失真衰减因子, 主要由于对像素的插值滤波引起, 取值大约为 0.85-1.0 之间^[1]。如果第 $n+1$ 帧参考前一帧编码, 则第 $n+1$ 帧的信道失真估计为

$$d_C(n+1) = ((1-p)\alpha\beta + p\beta)d_C(n) + p\bar{d}_Y \quad (8)$$

3.3 编码获益

当无传输差错时, 编/解码器的重建帧一致, 第 n 帧的传输失真等于信源失真。当有传输差错时, 如果采用容错视频编码技术, 将导致编码失真增加, 增加量为 $\Delta d_S(n)$, 同时也不可避免的产生信道失真 $d_C(n)$, 与无传输差错情况相比较, 有传输差错时的传输失真增大。定义传输失真的平均增加量为信道代价均值 \bar{d}_E :

$$\bar{d}_E = \frac{1}{N} \sum_{i=1}^N (\Delta d_Q(i) + d_C(i)) \quad (9)$$

其中 N 为序列总帧数。

为了分析信道失真对视频的传输失真的影响, 先考虑当前帧的候选参考帧为前一帧和前一正确核心帧的情况。如图 4(a)所示, 假定第 n 帧的信道失真较小, 在编码第 n 帧后面几帧时, 由于参考正确核心帧编码的编码代价较大, 因此选择参考前一帧编码。

假定在第 $n+w$ 帧前的各帧都参考前一帧编码, 根据式(9)可知第 $n+s$ ($s \leq w$) 帧的信道失真估计为

$$d_C(n+s) = \gamma^s d_C(n) + \frac{1-\gamma^s}{1-\gamma} p\bar{d}_Y \quad (10)$$

其中 $\gamma = (1-p)\alpha\beta + p\beta$ 。

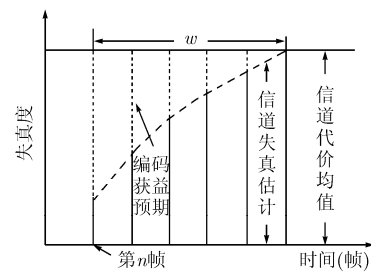
当参考前一帧编码时, 信道失真会不断增加。当信道失真达到一定程度时, 就需要参考前一正确核心帧编码来降低信道失真。因为不希望信道失真大于信道代价均值, 所以期望在信道失真约等于信道代价均值时选择参考前一正确核心帧编码, 即 $d_C(n+w) \approx \bar{d}_E$ 。 w 估计为

$$w = \left\lceil \log_{\gamma} \left[\frac{\bar{d}_E - p\bar{d}_Y / (1-\gamma)}{d_C - p\bar{d}_Y / (1-\gamma)} \right] \right\rceil \quad (11)$$

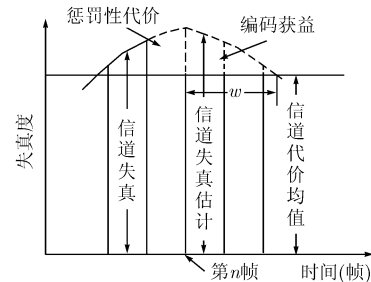
w 为信道失真增加到信道代价均值前, 帧间编码帧数的估计值, 称作扩散距离。相对于信道代价均值的编码获益 $d_B(n)$, 定义为各帧信道代价均值减去信道失真之差在扩散距离内的总和, 即

$$d_B(n) = \sum_{s=0}^w (\bar{d}_E - d_C(n+s)) \quad (12)$$

虽然期望在信道失真约等于信道代价均值时选择参考正确核心帧。但是对于某些帧, 在信道失真接近信道失真代价均值时, 因为参考正确核心帧的编码代价很大, 所以继续选择参考前一帧, 导致信道失真大于信道代价均值。如果此时继续选择参考前一帧, 信道失真将更大, 因此增大了参考正确核心帧的迫切性。为了反映这种迫切性, 这里给予继续参考前一帧编码一个惩罚性代价。如图 4(b), 第 n 帧的惩罚性代价大小为 $\sum_{s=-2}^0 (d_C(n+s) - \bar{d}_E)$ 。显然编码获益应该等于惩罚性代价的负数。为了更好地从数学期望角度反映编码获益, 编码获益不直接取惩罚性代价的负数, 而是采用对称方法, 假定第 n 帧后各帧的掩盖失真等于 $-\bar{d}_Y$, 使信道失真估计值按对称形式降低到信道代价均值, 从而编码获益定义保持与式(12)相同, 而其中的 $d_C(n+s)$ 和 w 分别估计为



(a) 信道失真小于信道代价均值



(b) 信道失真大于信道代价均值

图 4 编码获益示意图

$$d_C(n+s) = \frac{d_C(n)}{\gamma^s} + \frac{(1-\gamma^s)p\bar{d}_Y}{\gamma^s - \gamma^{s-1}}, \quad d_C(n) > \bar{d}_E \quad (13)$$

$$w = \left\lceil \log_{\frac{1}{\gamma}} \left[\frac{\bar{d}_E - \gamma p \bar{d}_Y / (1-\gamma)}{\bar{d}_C - \gamma p \bar{d}_Y / (1-\gamma)} \right] \right\rceil \quad (14)$$

当考虑多个可选参考帧时, 编码获益定义没有任何变化。因为参考不同参考帧编码的信道失真不同, 所以编码获益也不同。两种编码模式的编码获益之差, 就等于这两种编码模式对视频信道失真影响的差异。如图5所示, 图中①和③所示区域正是参考正确核心帧与参考前一帧对信道失真影响的差异, 而②所示区域为参考前一核心帧相对于参考前一帧对传输失真影响的差异。

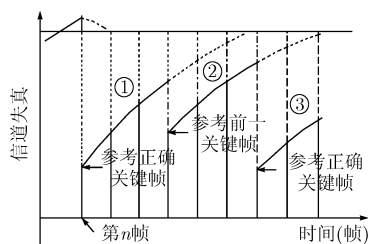


图5 多参考帧预编码时的宏块

相对与参考第 r' 帧编码, 参考第 r 帧编码时, 视频序列的信道失真降低的期望值 $\Delta d_B(r, r')$ 为

$$\Delta d_B(r, r') = \bar{d}_B(n, r) - \bar{d}_B(n, r') \quad (15)$$

3.4 参考帧选择

如果参考第 r 帧比参考第 r' 帧编码更优, 则参考第 r 帧编码能降低视频传输失真, 即下面不等式成立:

$$\begin{aligned} \Delta d_B(r, r') - \Delta d_Q(r, r') > 0 \\ \Rightarrow \bar{d}_B(n, r) - \phi \times R(n, r) > \bar{d}_B(n, r') - \phi \times R(n, r') \end{aligned} \quad (16)$$

根据式(16), 可得最佳参考帧 r^{\otimes} 为

$$\begin{aligned} r^{\otimes} &= \arg \min_{r \in \text{REF}} (\bar{d}_B(n, r) - \phi R(n, r)) \\ &= \arg \min_{r \in \text{REF}} (\bar{d}_B(n, r) - \phi \times (1 + k_Q \log_2(n-r)) B_p) \end{aligned} \quad (17)$$

其中 REF 为候选参考帧的集合, 在 ARPS 算法中包含前一帧、前一核心帧和前一正确核心帧。

4 模拟实验

模拟实验在 H.264 的测试模型 JM9.8 上进行, 模拟仿真采用文献[7]实验4的 GOB 交织打包传输技术, 差错掩盖假定为时域替代, 即对受损宏块采用前一帧相同位置的宏块替代。采用 IPPP 的编码模式编码 128 帧, 帧率为 15 帧/秒。采用 Internet 的数据包丢失模板文件 3, 5, 10 和 20, 其实际丢率分别为 3.3%, 5.6%, 11.5 和 20.8%^[5]。

表1为不同RPS方法在“Carphone”序列的亮度信号的PSNR性能比较, 实验设置丢包率设为5.6%。

图6显示了不同RPS方法在不同丢包率下的亮度信号的PSNR性能, 实验设置编码码率为64kbps, 最大往返时延为

表1 不同RPS方法在“Carphone”中的PSNR性能比较

RPS 方法	最大往返时延(ms)	编码码率(kbps)			
		64	128	196	256
ARPS	200	32.27	34.46	35.69	36.46
ACK	200	32.12	33.79	34.5	34.86
NACK	200	31.98	34.39	35.67	36.45
ARPS	400	32.53	33.87	35.20	36.04
ACK	400	31.59	33.66	35.06	35.97
NACK	400	31.21	32.54	33.10	33.39
ARPS	800	31.59	33.71	34.78	35.58
ACK	800	3.56	33.09	34.56	35.55
NACK	800	30.64	31.85	32.25	32.52

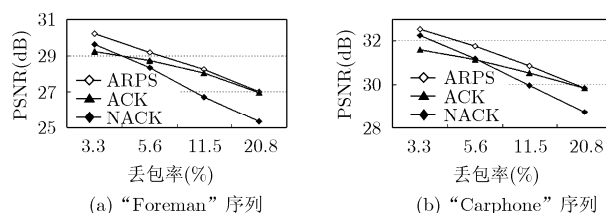


图6 ARPS在不同丢包率下的性能

400ms。图7显示了ARPS在不同编码码率下的亮度信号的PSNR性能。实验表明, 在高编码码率和高丢包率时, ARPS与ACK模式性能相当, PSNR性能提高不明显, 一般提高不超过0.3dB; 当与NACK模式相比, PSNR性能有明显提高, 提高在1dB以上。在低丢包率和低编码码率时, ARPS比NACK模式的PSNR性能相比提高较小, 一般为0.2~0.4dB; 但是比ACK模式的PSNR性能提高较大, 提高在1dB左右。在丢包率中等时, ACK模式和NACK模式的性能接近, ARPS均优于这两种模式, PSNR性能提升约0.5~1.0 dB。同时当往返时延的间隔帧数增大时, ARPS比RPS的ACK模式和NACK模式的性能优势更为明显。

5 结束语

经过对经典RPS不足之处的分析, 提出了基于率失真优化的自适应参考帧选择算法。编码器利用解码器的反馈信息准确估计信道失真, 实现了编码代价和编码获益的精确计算。在联合信源信道的率失真优化框架内选择参考帧, 实现了信源编码和信道编码的最佳平衡, 提高了视频在不同特征网络中传输的鲁棒性, 改善了视频传输质量。在不同丢包率

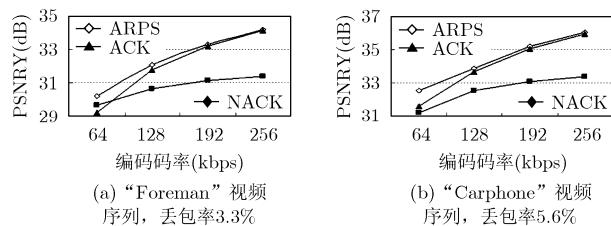


图7 ARPS在不同编码码率下的性能

下, ARPS 都具有比经典 RPS 的两种模式更优的 PSNR 性能。

参 考 文 献

- [1] He Zhi-hai, Cai Jian-fei, and Chen Chang-wen. Joint source channel rate-distortion analysis for adaptive mode selection and rate control in wireless video coding. *IEEE Trans. on Circuits and Syst. For Video Technol.*, 2002, 2(6): 511-523.
 - [2] Fukunaga S, Matsumura Y, and Nakai T. Error resilient video coding controlled by backward channel signaling. *Signal Processing: Image Communication*, 1999, 14(5): 531-540.
 - [3] Liang Y J, Setton E, and Girod B. Channel-adaptive video streaming using packet path diversity and rate-distortion optimized reference picture selection. Proc. IEEE Fifth Workshop on Multimedia Signal Processing, Virgin Island, Dec. 2002: 420-423.
 - [4] Lin Shu-nan, Mao Shi-wen, and Wang Yao, *et al.*. A reference picture selection scheme for video transmission over ad-hoc networks using multiple paths. IEEE International Conference on Multimedia and Expo2001, Tokyo Japan, 2001: 97-100.
 - [5] Aggelos K. Katsa Ggelos, Yifta Eisenberg, and Fan Zhai, *et al.*. Advances in efficient resource allocation for packet-based real-time video transmission. *Proc. IEEE*, 2005, 93(1): 135-147.
 - [6] Zhang Yuan, Gao Wen, and Lu Yan, *et al.*. Joint source-channel rate-distortion optimization for H.264 video coding over error-prone networks. *IEEE Trans. on Multimedia*, 2007, 9(3): 445-454.
 - [7] Wenger Stephan. H.264/AVC Over IP[J]. *IEEE Trans. on Circuits and System for Video Technology*, 2003, 13(7): 645-656.
- 彭强: 男, 1962年生, 教授, 博士生导师, 主要从事图形/图像处理、视频图像编码与传输、虚拟现实与多媒体技术、智能交通等领域的研究工作。
- 杨天武: 男, 1971年生, 博士生, 研究方向为数字视频编码与传输、机器人足球等。
- 张庆明: 男, 1975年生, 博士生, 研究方向为数字视频编码与传输。