

基于空间相关性的复杂度可分级帧间模式选择^{*})

冯 镛 朱光喜 刘文予

(华中科技大学 武汉光电国家实验室 武汉 430074)

摘要 本文提出了一种针对 H. 264 的复杂度可分级的帧间模式选择算法,该算法采用一张概率表将当前宏块的 7 种预测模式与相邻的上侧和左侧宏块的最优模式联系起来。对于每一种相邻宏块模式的组合情况,该表将当前宏块的 7 种预测模式按照出现概率的高低进行排序,从而可以确定最可能模式、次最可能模式等。在此基础上,调整候选模式数量以匹配终端平台不同的复杂度约束水平,同时又可以最大可能地保证最优模式处在所选择的部分模式子集中。该算法具有 4 个复杂度级别:40%,54%,70%和 82%,实验证明该方法可以灵活有效地降低编码器复杂度,同时编码质量下降很小,不会影响主观效果。

关键词 视频编码,帧间预测,空间相关性,复杂度可分级,H. 264

Complexity Scalable Inter Modes Decision Algorithm for H. 264 Based on Spatial Correlation

FENG Bin ZHU Guang-Xi LIU Wen-Yu

(Wuhan Optical-Electronic National Laboratory, Huazhong Univ. of Sci. & Tech. Wuhan 430074)

Abstract In this paper, the complexity scalability for H. 264 motion estimation was explored. The complexity scalable algorithm proposed uses a probability table to relate the prediction modes of adjacent MBs to that of the current MB. For each combination of adjacent MBs, the probability table includes a list of prediction modes in order of expected occurrence. Based on this table, an ordering of the most probable, next most probable prediction mode etc. is determined. Hence the number of modes used in inter prediction can be adjusted under different complexity constraints while the optimum mode is guaranteed to lie in the chosen modes with great possibility. This proposed algorithm has 4 computation scales: 40%, 54%, 70%, and 82%. The experimental results show that the proposed method can save the encoding time effectively and flexibly with acceptable performance degradation.

Keywords Video coding, Inter prediction, Spatial correlation, Complexity scalable, H. 264

1 引言

H. 264^[1]是由联合视频小组 JVT 发布的最新的视频编码标准。它采用了许多先进的编码技术,包括可变块大小运动补偿 1/4 像素精度的运动补偿以及多参考帧等等。这些技术使得 H. 264 的编码性能大大优于当前的其他编码标准。例如,和 MPEG2 相比,H. 264 在相同视觉质量下可以节省近 50%的码率^[2]。但是,这样优异的编码性能是以计算复杂度的增加为代价的。H. 264 的编码复杂度要比 MPEG4 的简单级别高 16 倍以上^[3],实验结果表明,运动估计与补偿是计算量最大的模块之一,大约占整个编码器的 1/3 左右,而率失真优化过程则需要消耗编码器 50%左右的计算量^[4]。这种方法适用于高端设备,因为此时如何获得更高的编码效率是首先要考虑的问题,而对于低端设备而言,由此带来的巨大复杂度则是难以承受的。

因此,一直有许多学者致力于快速运动估计算法的研究。文[5]中提出了一种通过不同模式间率失真代价值的预测关系来选择最优帧间编码模式的算法,但是这种预测关系很难准确得到。Wu D 等人利用视频对象在空域的一致性以及时域的平稳性来减少帧间预测中候选模式的个数^[6]。Ting 等人提出了一种中心偏移的参考帧预选择算法,目的是减少运动估计中需要遍历的参考帧个数^[7]。这些算法都可以在编码

质量降低不大的情况下,有效地减少编码器的计算复杂度,但是算法的灵活性较差,无法适应从高端设备到低端设备的不同计算复杂度要求。

为了解决这一问题,本文提出了一种复杂度可分级的运动估计算法,该算法具有 4 个不同的复杂度级别,复杂度越高,编码质量越好。本文算法采用一张概率表将当前宏块的 7 种预测模式与相邻的上侧和左侧宏块的最优模式联系起来。对于每一种相邻宏块模式的组合情况,该表将当前宏块的 7 种预测模式按照出现概率的高低进行排序,从而可以确定最可能模式、次最可能模式等。在此基础上,调整用于帧间预测的模式数量以匹配终端平台不同的计算资源,同时又可以最大可能的保证最优模式处在所选择的部分模式子集中。

本文在第 2 节中简要介绍 H. 264 中帧间模式选择的方法;第 3 节详细描述并分析提出的复杂度可分级算法;实验结果和相关分析将在第 4 节中给出;最后对全文进行总结。

2 H. 264 帧间模式选择算法

在以前的视频编码标准中,运动估计的最小单位是一个宏块(16×16),这意味着宏块内所有的像素点都具有相同的运动矢量。这种方法适用于处于静止宏块或一致性运动的宏块。但是,如果宏块内包含有丰富的细节,或者宏块处在不同

^{*}基金项目:国家自然科学基金资助项目(60372067),高等学校博士点基金资助项目(20040487009)。冯 镛 博士研究生。

物体的边界上,这时只采用一个运动适量进行运动补偿是无法准确描述物体的运动的,从而会造成较大质量的下降。H.264 对于帧间宏块定义了 7 种编码模式,包括 SKIP 模式、4 种帧间模式以及 2 种帧内预测模式,图 1 画出了所有可能的编码模式示意图。一个帧间编码宏块可以以 4 种尺寸大小进行分割,分别是 16×16、16×8、8×16 以及 8×8。如果采用 8×8 宏块模式编码,则对于每一个 8×8 子块,还可以进一步以 8×8、8×4、4×8 以及 4×4 进行子宏块分割。模式为 0 的宏块采用 SKIP 模式编码,而模式 9 和 10 则表示两种帧内编码模式。为了确定每个宏块的最优编码模式,编码器会计算所有可能模式对应的率失真代价,选择代价最小的模式对宏块进行编码。率失真代价函数如(1)式所示:

$$J = D + \lambda * R \quad (1)$$

其中 J 是代价函数, λ 是用于模式选择的拉格朗日因子, R 是当前宏块编码的比特数, D 是原始宏块 M 和对应的重建宏块 M' 之间的均方误差和,具体定义如下:

$$D = \sum_{i=0}^{15} \sum_{j=0}^{15} (M(i,j) - M'(i,j))^2 \quad (2)$$

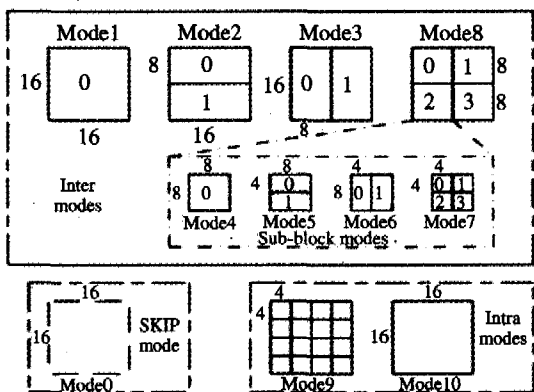


图 1 H.264 中帧间宏块的各种编码模式

3 复杂度可分级算法

运动估计中候选模式的数量会直接影响到编码器的计算复杂度,采用更多的模式,则需要消耗更多的计算功率。如果候选模式的数量能够随着不同的终端设备功率约束而进行灵活的调整,则帧间预测可以满足从高端设备到低端设备等不同应用平台的要求。同时,减少候选模式数量所带来的质量下降也是一个重要的问题。如果当候选模式减少时,最优模式能够以最大的概率属于所选择的模式子集,则编码质量的损失可以减少到最小。

H.264 标准所采用的运动估计所具有的一个特点就是,当物体的边界穿过一个宏块时,这个宏块将被分割成更小的块^[8]。如图 2 所示,如果物体的边界垂直穿越上侧的宏块,则当前宏块中也很有可能包含一个垂直边界;基于同样的理由,如果左侧宏块的最优模式为 16×8,这表示有水平边界穿越左侧宏块,则当前宏块的最优模式也有很大的概率是 16×8。因此,根据图像的空间相关性,可以得出以下的推论:最优帧间预测模式之间也是具有空间相关性的。表 1 给出了经过大量标准视频序列统计得到的当前宏块最优模式与相邻宏块模式之间的关系表。这里 $M(l)$ 和 $M(u)$ 分别表示左侧和上侧宏块的最优模式。从表 1 中可以看出,在大多数情况下,当前宏块的最优模式是相邻宏块模式中的一种。这也证明了最优模式之间是具有空间相关性的。

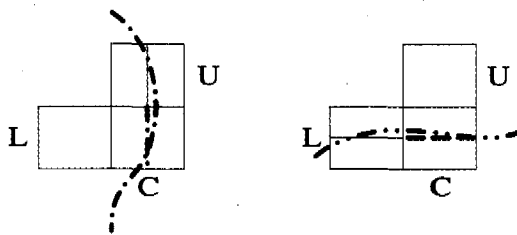


图 2 相邻宏块模式的相关性

表 1 不同相邻模式条件下当前宏块的最优模式

$M(l) \backslash M(u)$	0	1	2	3	8	9	10
0	0	0	0	0	0	0	0
1	0	1	1	1	8	9	1
2	0	1	1	1	8	9	1
3	0	1	1	1	8	9	1
8	0	8	8	8	8	8	8
9	0	1	8	8	8	9	9
10	0	0	0	0	8	9	10

更进一步,如果能够得到在每一种相邻模式组合情况下,当前宏块 7 种预测模式出现概率的优先级顺序,则可以更加灵活的进行运动估计。当左侧和上侧宏块模式已知时,可以根据终端平台不同的功率约束水平来调整运动估计中候选模式的个数,选择最有可能的几个模式进行预测,如最可能模式、次最可能模式等等。因此,本文算法采用一张概率表将当前宏块的 7 种预测模式与相邻的上侧和左侧宏块的最优模式联系起来。对于每一种相邻宏块模式的组合情况,该表将当前宏块的 7 种预测模式按照出现概率的高低进行排序,从而可以确定最可能模式、次最可能模式等。在此基础上,调整用于帧间预测的模式数量以匹配终端平台不同的计算资源,同时又可以最大可能的保证最优模式处在所选择的部分模式子集中。令 l 和 u 分别为左侧和上侧宏块的模式索引, $M(l)$ 和 $M(u)$ 表示的是模式集合 $M = \{0, 1, 2, 3, 8, 9, 10\}$ 中的一种;令 p 表示在某一左侧和上侧宏块模式组合条件下, M 中的 7 种预测模式在当前宏块中按照出现概率从大到小进行排序后的索引号;显然 l, u 和 p 取值范围均为 $[0:6]$ 。通过大样本统计的方法建立在左侧及上侧宏块模式组合条件下,当前宏块模式的优先级关系表 $Correlation[7][7][7]$ 。其中 $Correlation[l][u][p]$ 表示的是当左侧宏块模式为 $M(l)$, 上侧宏块模式为 $M(u)$ 时,当前宏块第 p 个最有可能出现的模式。

表 2 中给出了具体的优先级关系表 $Correlation[7][7][7]$,在此基础上可以选择最有可能的几种模式进行帧间预测。例如,当左侧宏块的模式为 1, 上侧宏块的模式为 3 时,表 2 中给出模式优先级顺序为 1、0、8、3、2、9 和 10。这说明对于当前宏块,模式 1 为最有可能出现的模式,模式 0 是次最可能出现的模式,以次类推。如果这时终端的功率约束水平允许可用模式数量为 3,则用前 3 个最有可能出现的模式来进行运动估计,即 1、0 和 8;如果允许的可用模式数量为 5 时,则采用的模式为 1、0、8、3 以及 2。采用更多的预测模式可以以更多的计算复杂度为代价,获得更好的编码质量。为了防止差错蔓延所带来大幅度质量下降,至少要采用 3 个模式来进行帧间预测。也就是说,本算法可以通过调整模式数量从 3 到 6 变化,来提供 4 个复杂度级别。当采用的模式数量为 7 时,最优预测模式等同于全模式搜索的结果,则不会有质量损

表 2 基于空间相关性的当前宏块模式优先级

M(l)\M(u)	0	1	2	3	8	9	10
0	0,1,8,3,2,10,9	0,1,3,2,8,10,9	0,1,8,3,2,9,10	0,1,3,8,2,9,10	0,8,1,3,2,9,10	0,1,9,8,2,3,10	0,1,10,3,2,9,8
1	0,1,8,3,2,10,9	1,0,8,2,3,9,10	1,0,8,2,3,9,10	1,0,8,3,2,9,10	8,1,3,0,2,9,10	9,1,8,3,2,0,10	1,0,2,3,8,10,9
2	0,1,2,8,3,10,9	1,8,2,0,3,9,10	1,8,2,0,3,9,10	1,8,3,0,2,9,10	8,1,2,3,0,9,10	9,8,1,2,3,0,10	1,2,0,8,10,9,3
3	0,1,8,3,2,9,10	1,0,8,3,2,9,10	1,0,8,2,3,9,10	1,8,3,0,2,9,10	8,1,3,0,2,9,10	9,8,1,3,2,10,0	1,3,0,8,2,9,10
8	0,8,1,2,3,9,10	8,1,2,0,3,9,10	8,1,2,0,3,9,10	8,1,3,2,0,9,10	8,1,3,2,0,9,10	8,9,1,2,3,10,0	8,1,3,2,9,0,10
9	0,1,8,2,9,3,10	1,8,9,2,3,0,10	8,9,1,2,3,0,10	8,9,1,3,2,0,10	8,9,1,3,2,0,10	9,8,1,3,10,2,0	9,10,1,3,2,8,0
10	0,1,10,2,3,9,8	0,1,10,2,3,8,9	0,1,10,8,2,9,3	0,1,3,8,9,10,2	8,1,3,0,9,2,10	9,10,1,0,3,8,2	10,0,1,9,2,3,8

失,同样也不会有复杂度的减少。

需要指出的是,对于每一帧图像中的第一行和第一列宏块,分别缺少上侧和左侧宏块模式,所以此时会采用所有的 7 种模式进行预测,这可以保证当前宏块最优模式的正确选择,同时使后续宏块不会有严重的差错蔓延。

4 实验结果

本文算法在 H. 264 编解码器 JM85^[9] 基础上实现,采用四个标准视频测试序列的前 100 帧进行测试,分别是“Bus”、“Mobile”、“Paris”、“Tempete”;图像大小为 QCIF (176 × 144);帧率为 30fps;编码帧结构为 IPPP;打开哈达码变换和率失真优化,采用统一变长编码(UVLC)进行熵编码。统计每个序列在 QP=28,32,36,40 四个点的亮度 PSNR 值和码率,并通过 AVSNR 软件计算平均编码增益。实验环境是包括 Celeron 1.7GHz CPU 和 384M 内存的个人 PC。

表 3 中列出的是本文算法和 JM85 的性能比较结果,负值的 Δ PSNR 和正值的 Δ Bitrate 表示本文算法性能相对 JM85 性能有所下降,这是由于该算法选用了较少的候选模式造成的。从表 3 可以看出,本文算法的编码质量随着采用的模式数量的增加而线性提高。当采用的模式数量为 3 时,性能损失最大,最大为 -0.13dB,平均的质量下降约为 -0.1dB。从人眼视觉的生理特性来看,当图像质量的下降在 -0.5dB 以内时,人的主观感觉并不会太大的变化。所以,本文方法造成的图像质量下降,因其幅度很小,基本上不会影响视频的主观视觉效果。图 3 给出了本文算法在模式数量为 3 时与 JM85 的性能比较曲线,从图中可以看出,这时本文算法已经具有足够好的编码性能。当功率约束水平允许时,可以采用更多的模式,这时的编码性能和 JM85 几乎相当。例如当模式数量为 6 时,性能下降仅为 -0.01dB,几乎可以忽略。

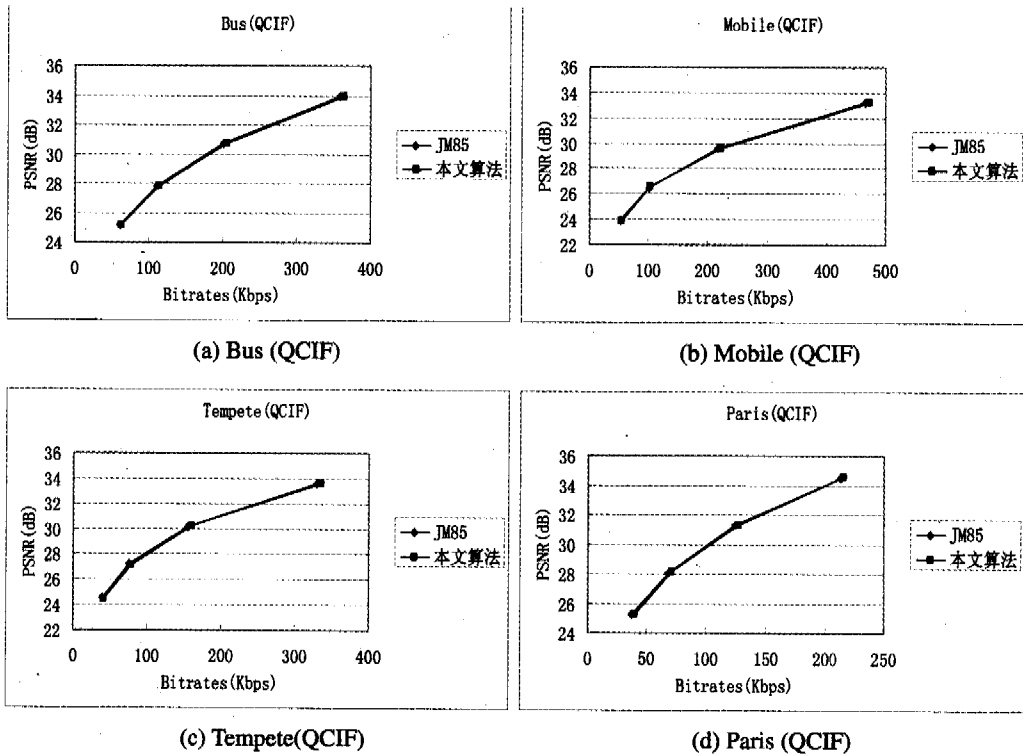


图 3 本文算法与 JM85 性能比较曲线(PSNR vs. Bitrates)

预测模式的减少可以相应的减少编码复杂度。表 3 中还比较了本文算法和 JM85 帧间编码时间,从表中可以看出,本文算法有 4 个复杂度级别,分别是 40%、54%、70% 以及 82%。需要指出的是,该时间数据是未经优化的测试软件在 PC 机上测试得到的,其中还包括比较耗时的磁盘读写等操作。所以在实际的通信终端上,采用本文方法的时间节省比

例应该更高,效果会更明显。

结论 本文利用图像的空间相关性,提出了一种复杂度可分级的帧间模式选择算法。该算法采用一张概率表将当前宏块的 7 种预测模式与相邻的上侧和左侧宏块的最优模式联系起来。对于每一种相邻宏块模式的组合情况,该表将当前宏块的 7 种预测模式按照出现概率的高低进行排序,从而可

以确定最可能模式、次最可能模式等。在此基础上,调整用于帧间预测的模式数量以匹配终端平台不同的计算资源,同时又可以最大可能的保证最优模式处在所选择的部分模式子集中。本文算法具有4个不同的复杂度级别:40%、54%、70%以及82%,同时实验结果证明,编码质量的损失不大,不会影响主观视觉效果。

表3 编码性能比较结果

序列	Bus			Mobile		
	Δ PSNR /dB	Δ Bitrate /%	复杂度 /%	Δ PSNR /dB	Δ Bitrate /%	复杂度 /%
3模式	-0.09	+2.03	39.24	-0.09	+2.21	40.14
4模式	-0.06	+1.23	55.42	-0.05	+1.12	54.87
5模式	-0.04	+0.81	69.37	-0.03	+0.62	70.02
6模式	-0.02	+0.30	83.10	-0.01	+0.21	82.87
序列	Tempete			Paris		
	Δ PSNR /dB	Δ Bitrate /%	复杂度 /%	Δ PSNR /dB	Δ Bitrate /%	复杂度 /%
3模式	-0.13	+3.02	38.75	-0.12	+2.38	39.02
4模式	-0.06	+1.56%	53.28	-0.08	+1.90	54.74
5模式	-0.02	+0.54%	68.79	-0.05	+1.12	67.26
6模式	-0.01	+0.21%	82.18	-0.03	+0.53	82.46

参考文献

- 1 Information Technology—Coding of Audio-Visual Objects— Part 10; Advanced Video Coding. Final Draft International Standard, ISO/IEC FDIS 14 496-10
- 2 Wiegand T, Sullivan G J, Bjntegaard G, Luthra A. Overview of the H. 264/AVC video coding standard. IEEE Transactions on Circuits And Systems for video tecnology, JULY 2003,13(7)
- 3 Zhou Minhua. Evaluation and Simplification of H. 26L Baseline Coding Tools. ITU-T Q. 6/16, Doc. #JVT-B030, 2002
- 4 Liang Yongfang, He Zhihai, Ahmad I. Analysis and design of power constrained video encoder. Emerging Technologies; Frontiers of Mobile and Wireless Communication. In: Proceedings of IEEE 6th Circuits and Systems Symposium on, 2004,1:57~60
- 5 Yin P, Cheong H Y, Tourapis A N, Boyce J. Fast mode decision and motion estimation for JVT/H. 264. In: Proc. IEEE ICIP, 3, Sep. 2003, 3:853~856
- 6 Wu D, Pan F, Lim K P. et al. Fast Intermode Decision in H. 264/AVC Video Coding. IEEE Transactions on Circuits and Systems for Video Technology, July 2005, 15:953~958
- 7 Ting C M, Po L M, Cheung C H. Center-biased frame selection algorithms for fast multi-frame motion estimation in H. 264. In: Proceeding of 2003 IEEE International Conference on Neural Networks and Signal Processing, Dec. 2003. 1258~1261
- 8 Huang Y W, Hsieh B Y, Wang T C, et al. Analysis and reduce of reference frames for motion estimation in MPEG-4 AVC/JVT/H. 264. In: Proceeding of the Int. Conf. On Acoustics, Speech, & Signal Processing, Sept. 2003, 2:809~812
- 9 JVT reference software version 8. 5. http://bs.hhi.de/~suehring/tml/download/old_jm/

(上接第35页)

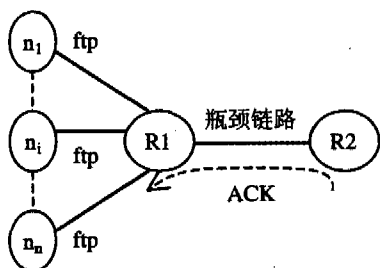


图3 仿真网络拓扑结构

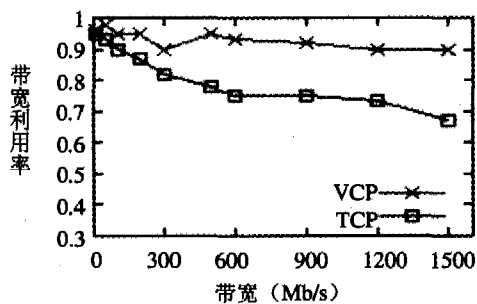


图4 不同带宽时的利用率

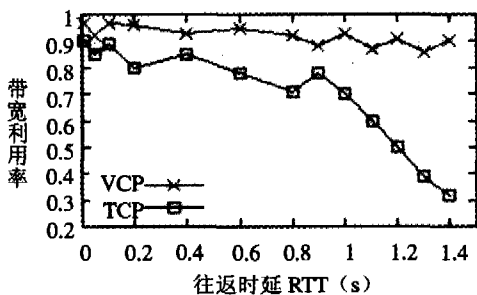


图5 不同RTT时的利用率

可以看出,当路由器R1、R2之间的链路带宽增加到1.5Gbs时,TCP的带宽利用率下降到70%以下,而VCP的带宽利用率则保持在90%以上,且随着带宽的继续增加,VCP始终保持较高的带宽利用率。同样如图5所示,当业务源的往返时延RTT增大为1s时,带宽利用率约高于60%,RTT继续增大时,带宽利用率严重下降。而RTT大于1s时,VCP的利用率保持在90%左右,虽然约有振荡,但一般都保持在80%以上。

讨论和进一步研究内容 VCP协议在体系结构上与TCP+AQM/ECN相同,都是路由器检测拥塞,源主机调整算法。但是VCP在路由器中采用负载因子来检测拥塞,而TCP是使用的队列。VCP使用了两个ECN位来通知拥塞,比TCP多了一位,但是却获得了性能的大幅度提升。VCP对现有的协议改动非常小,且可扩展性良好,有利于逐步的实施。进一步的研究内容包括:端到端VCP、与TCP的友好性即TCP-friendliness以及在无线链路上的应用等等。

参考文献

- 1 Jacobson V. Congestion avoidance and control. In: Proc. ACM SIGCOMM'88, August 1988
- 2 Lakshman T, Madhow U. The performance of TCP/IP for networks with High Bandwidth-delay products and Random Loss. IEEE Trans. On network, June 1997
- 3 Xia Y, Subramanian, Stoica I, Kalyanaraman S. One more bit is enough. In: Proc. ACM SIGCOMM'05, August 2005
- 4 Xia Y, Subramanian, Stoica I, Kalyanaraman S. One more bit is enough. [UC Berkeley Tech Report]. June 2005
- 5 Ramakrishnan K K, Floyd S. The Addition of Explicit Congestion Notification (ECN) to IP. IETF RFC 3168, September 2001
- 6 Katabi D, Handley M, Rohrs C. Congestion Control for High Bandwidth-Delay Product Networks. In: Proc. ACM SIGCOMM'02, August, 2002
- 7 Chiu D, Jain R. Analysis of the Increase/Decrease Algorithms for Congestion Avoidance in Computer Networks. J. of Computer Networks and ISDN, 1989,17(1):1~14
- 8 Yang Y, Lam S. General AIMD Congestion Control. ICNP'00, November 2000
- 9 Hollot C, Misra V, Towsley D, Gong W. On Designing Improved Controllers for AQM Routers Supporting TCP Flows. INFOCOMM'01, April 2001
- 10 The network simulator ns-2. <http://www.isi.edu/nsnam/ns>

VCP协议具有良好的瓶颈链路平均带宽利用率。从图4