

# JPEG2000 算术编码器的算法优化和 VLSI 设计

刘文松<sup>1</sup>, 朱 恩<sup>1</sup>, 王 健<sup>2</sup>, 徐龙涛<sup>1</sup>, 林 叶<sup>1</sup>

(1. 东南大学射频与光电集成电路研究所, 江苏南京 210096; 2. 中国科学院自动化研究所, 北京 100190)

**摘 要:** 研究了 JPEG2000 算术编码器的算法和电路实现. 提出了重归一化规程的一种新的顺序结构, 通过添加独立的总移位次数预测规程, 使得编码算法可以一次性顺序完成当前上下文的处理. 据此设计了具有从流水线的三级流水线电路结构, 流水线用于处理无编码字节输出的常规情况, 从流水线单独处理编码字节的输出, 从而有效缩短了各级电路的关键路径延时. 基于 TSMC 0.18 $\mu\text{m}$  工艺的综合结果表明, 系统时钟频率为 578MHz, 吞吐量约为 520 Msymbols/s. 与采用相同工艺的国内外研究成果相比, 分别提升了 40% 和 26% 以上.

**关键词:** JPEG2000; 算术编码; 重归一化规程; 流水线结构

**中图分类号:** TN492      **文献标识码:** A      **文章编号:** 0372-2112 (2011) 11-2486-06

## Optimization Algorithm and VLSI Design of the Arithmetic Coder in JPEG2000

LIU Wen-song<sup>1</sup>, ZHU En<sup>1</sup>, WANG Jian<sup>2</sup>, XU Long-tao<sup>1</sup>, LIN Ye<sup>1</sup>

(1. Institute of RF and OE-ICs, Southeast University, Nanjing, Jiangsu 210096, China;

2. Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China)

**Abstract:** The algorithm and the hardware implementation of Arithmetic Coder (AC) in JPEG2000 are studied. A new kind of the sequence structure of the Renormalization Procedure (RP) is proposed. With the independent total-shift prediction procedure added, the current context can be processed serially by AC without loop computation. Based on the proposed algorithm, the 3-stage pipeline architecture with a slave pipeline is designed, where the pipeline is used to process the common situation of no byte-output and the slave pipeline is used to detect and process the byte-output situation separately, in order to reduce the critical path of the sub circuits. The synthesis result by the technology library of TSMC 0.18 $\mu\text{m}$  shows that the system clock frequency is 578MHz and the throughput is about 520 Msymbols/s. Comparing with the published works by the same technology library at home and abroad, they are optimized by 40% and 26% at least respectively.

**Key words:** JPEG2000; AC; renormalization procedure; pipeline architecture

## 1 引言

随着各类高清数字图像的应用普及, 对图像编码器的实时处理能力提出了更高要求. 国际最新的图像压缩标准 JPEG2000 具备帧内编码、压缩效果好、支持感兴趣区域编码和渐进式传输等品质, 因而在高端数字图像应用中得到了推广<sup>[1]</sup>. 在 JPEG2000 编码过程中, 首先采用离散小波 (Discrete Wavelet Transform, DWT) 处理原始图像, 之后采用位平面编码器 (Bit Plane Coder, BPC) 处理小波系数子带生成上下文和码值 (Context/Decision, CX/D), 再送入算术编码器 (Arithmetic Coder, AC) 进行压缩编码<sup>[1~3]</sup>. 其中, AC 每读入一个 CX/D, 都要更新符号概率区间, 并在必要时采用循环迭代计算使其处于约定范围内, 之后才可以读入新的数据. 循环迭代计算涉及复杂的数据控制和组合逻辑判断, 不宜于流水线实现, 从

而影响了系统整体性能. 如何优化算法, 设计简洁高效的电路结构, 成为国内外学者研究的热点之一.

该领域有代表性的研究成果有: 文献[4~7]研究了多 CX/D 并行输入编码的方案. CX/D 并行输入编码充分发挥了电路并行的优势, 提高了系统吞吐量. 但从已发表文献来看, 并行编码方案的引入导致了编码关键路径较长, 时钟频率较低; 且必须与 BPC 的输出接口紧耦合, 对 CX/D 的读入进行有效控制. 文献[4]最早提出了主零点预测的思想, 即通过分析概率估值的主零点分布规律, 以查表形式求得重归一化规程 (Renormalization Procedure, RP) 期间所需的总移位次数. 在此基础上文献[8]和[9]着重对重归一化规程进行了研究, 提出了可以避免循环迭代计算的顺序算法结构. 但是, 文献[8]对算术编码算法的整体结构改动较大, 且判决条件分支过多, 交叉进行, 不适于硬件实现. 而文献[9]仅仅考虑



对应的  $Index$  和  $MPSS$ .  $AC$  每收到一个上下文时, 从  $PET$  中取出前述参数, 按照图 2 所示流程对  $A$ 、 $C$  和  $CT$  操作, 并在结束时更新  $Index$  和  $MPSS$ . 如此循环, 直到处理完所有的  $CX/D$ .

## 2.2 顺序结构的 AC 算法

文献[4]最早提出了主零点预测的思想, 并将其应用在多  $CXD$  并行输入处理电路设计中. 在此基础上, 文献[8]和[9]进一步发展细化, 将其应用在了  $RP$  算法顺序结构设计中. 主零点预测的基本原理是这样的:  $RP$  循环执行的条件是逐次移位  $A$ , 直至使其处于约定区间内.  $A$  在新的  $CX/D$  输入时, 通过式(1)和式(2)来更新. 显然, 通过式(2)更新时必定执行  $RP$ , 而  $A$  在  $RP$  执行期间的总移位次数  $TS$  即等于  $Q_e$  的主零点个数  $LZ$ . 依据主零点个数对  $PET$  中  $Q_e$  进行分类, 如表 1 所示. 在实际使用时, 可根据表 1 扩充  $JPEG2000$  的  $PET$ , 从而以查表方式获得.

表 1 概率估值与主零点的关系

主零点	概率估值 $Q_e$		主零点	概率估值 $Q_e$	
	上边界	下边界		上边界	下边界
1	0x7FFF	0x4000	9	0x007F	0x0040
2	0x3FFF	0x2000	10	0x003F	0x0020
3	0x1FFF	0x1000	11	0x001F	0x0010
4	0x0FFF	0x0800	12	0x000F	0x0008
5	0x07FF	0x0400	13	0x0007	0x0004
6	0x03FF	0x0200	14	0x0003	0x0002
7	0x01FF	0x0100	15	0x0001	0x0001
8	0x00FF	0x0080			

当通过式(1)更新时, 又分为两种情况: (1) 当  $A$  更新后已然处于  $[0.75, 1.5]$  区间内, 无需执行  $RP$ , 或认为总移位次数为 0; (2) 当更新结果小于 0.75 时, 需要执行  $RP$ . 已知  $A$  的最小值为  $0x8000$ ,  $Q_e$  的最大值<sup>[1]</sup> 为  $0x5601$ , 则  $A$  更新后的最小值为  $0x29ff$ . 结合表 1, 可通过与  $0x4000$  比较来进一步确定总移位次数是 1 还是 2. 结合以上讨论, 完整的总移位次数预测规程如图 3 所示.

$RP$  算法中涉及  $A$ 、 $C$  和  $CT$  的更新. 其中,  $C$  和  $CT$  的更新也取决于  $A$  的移位次数. 因此, 在执行  $RP$  前预测到  $A$  所需的总移位次数, 就为一次性顺序的处理完  $A$ 、 $C$  和  $CT$  奠定了基础. 分析总移位次数预测算法, 知  $Q_e$  等于  $0x0001$  时, 总移位次数最大为 15. 同时由图 1 知,  $CT$  的递减次数等于  $A$  的移位次数. 当  $CT$  归零时,  $RP$  调用字节输出规程 (Byteout Procedure, BP) 输出 1 字节的编码, 并将  $CT$  复位为 7 或 8. 并已知  $RP$  的执行条件是  $A$  完成总的移位次数. 结合以上条件可以推出: 每执行一次  $RP$ , 仅存在无字节输出、单字节输出和双字节输出三种情况. 因此, 若在执行  $RP$  前通过总移位次数

预测规程确定处理当前上下文所需的总移位次数, 便可直接按照上述三种情况对  $A$ 、 $C$  和  $CT$  一次性完成更新. 新的  $RP$  顺序结构如图 4 所示.

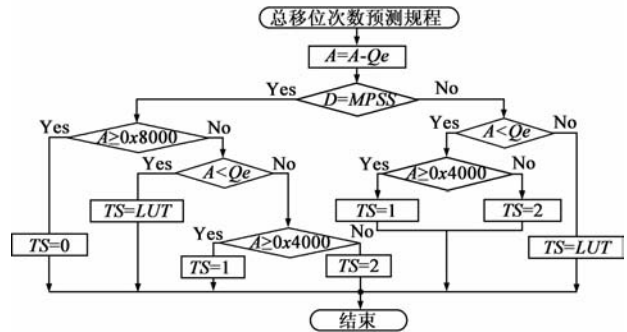


图 3 总移位次数预测规程图

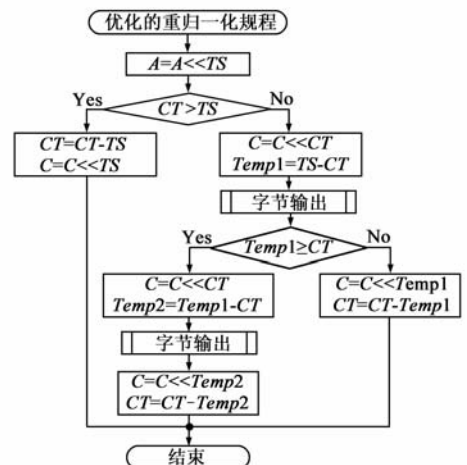


图 4 优化的顺序结构的重归一化规程

(1) 无字节输出: 当  $TS$  小于  $CT$  时, 仅需用  $TS$  对  $A$ 、 $C$  和  $CT$  更新,  $RP$  即可结束; (2) 单字节输出: 当  $TS$  大于等于  $CT$  时, 必有字节输出. 首先使用  $TS$  对  $A$  一次性更新, 并求得执行本次  $RP$  所需的剩余移位步数  $Temp1$ . 用  $CT$  对  $C$  进行更新后, 调用 BP 输出第一个字节, 期间复位  $CT$ . 进一步比较  $Temp1$  与  $CT$ . 若  $Temp1$  小于  $CT$ , 属于单字节输出情况, 用  $Temp1$  更新  $C$  和  $CT$  后退出  $RP$  即可; 否则属于双字节输出情况; (3) 双字节输出: 用  $Temp1$  与  $CT$  求得执行本次  $RP$  所需的剩余移位步数  $Temp2$ . 用  $Temp1$  更新  $C$  后, 调用 BP 输出一个字节, 期间复位  $CT$ ; 用  $Temp2$  对  $C$  和  $CT$  进行最后的更新,  $RP$  结束.

新的  $RP$  顺序结构使得  $AC$  可以一次性顺序的处理完当前上下文, 无需迭代循环计算. 相对于文献[8]和[9], 通过添加了独立的总移位次数预测规程, 从而可以无缝的替代原有部分; 无需改动算术编码算法的主体结构; 逻辑分支间不存在交织判断; 同时囊括了无字节输出、单字节输出和双字节输出三种情况, 保证了编码结果的正确. 适于硬件实现.

### 3 基于优化算法的电路结构

#### 3.1 结构分析

JPEG2000 算术编码过程中, 88% 甚至更多的 CX/D 在输入算术编码器后不会生成编码字节; 约 10% 左右的 CX/D 会生成 1 字节编码; 仅有约 0.0002% - 0.0006% 的 CX/D 会生成 2 字节编码<sup>[14]</sup>. 因此, 如果能在保证电路关键路径的同时, 设计常规的流水线电路处理无字节输出情况, 设计单独电路预测字节输出情况并加以处理, 将极大提高系统吞吐量. 新 RP 算法的顺序结构不仅可以有效降低计算量和判断逻辑, 更使得算术编码器可以一次性顺序的处理完成当前上下文. 因此可以根据概率估值更新的关联条件, 将算法分解到各级流水线实现. 由于只需比较更新后的 TS 与 CT 即可判断出字节输出情况, 因此无需设计单独的检测电路. 字节输出判断是逐次进行的, 即在判断完第一个字节输出后用相同的方法继续判断是否输出第二个字节, 因此可以设计从流水线, 仅在需要输出字节时才启动.

基于这种思想设计带有从流水线的三级流水线电路结构, 其结构框图如图 5 所示, 包含 6 个主要部分: 上下文状态存储和求取  $Q_e$ , Switch 等参数的概率估值查表电路, 构成第一级流水线; A, TS, MPSS 和 Index 更新电路, 构成第二级流水线; C 与 CT 更新电路, 构成第三级流水线; 字节输出电路, 构成从流水线; 用于存储 19 种上下文对应的 Index 和 MPSS 的上下文状态存储电路以及控制电路.

三级主流线用于处理发生概率最高的无字节输出的情况. 从流水线仅在检测到字节输出时启动. 完整的流水线运行流程示意图如图 6 所示. 第一级流水线读

取 CX/D, 先从上下文状态存储电路中取出对应的 MPSS 与 Index, 再查表求取  $Q_e$ , Switch 等参数. 第二级流水线根据这些概率参数更新 A, TS, MPSS 和 Index, 并将当前上下文的 MPSS 和 Index 存回上下文状态存储电路; 第三级流水线同样根据这些概率参数更新 C 和 CT, 并比较更新后的 TS 和 CT. 当没有编码字节输出时, 流水线将这样持续运行. 当检测到有编码字节需要输出时, 控制模块停止读入新的 CX/D, 启动从流水线: 即在下一个周期开始时将第三级流水线更新后的 C 和 CT 传递给字节输出电路, 输出一个编码字节. 字节输出电路将更新后的 C 和 CT 传回控制电路, 继续判断是否有第二字节输出. 如果是则继续启动从流水线输出编码字节, 否则启动常规流水线处理新的 CX/D.

注意到, 第一级流水线读入新的 CX/D 后从上下文状态存储电路中取出概率估值, 而第二级流水线则是计算前一 CX/D 的 A, Index 和 MPSS, 在下一个时钟周期上升沿存入上下文状态存储电路. 已知算术编码算法根据当前上下文的 Index 和 MPSS 求取概率估值参数, 然后计算 A, Index 和 MPSS. 因此, 当收到的前后两个上下文不一致时, 这样操作不会产生逻辑冲突. 但当连续收到同一个上下文时, 第二级流水线尚未将更新后的 Index 和 MPSS 返回上下文状态存储电路, 导致第一级流水线会读取错误的概率估值参数. 为此, 设计了状态忽略(State Neglect, SN) 技术来克服该问题. 其基本思想是: 在时钟周期开始时, 检测是否连续读入了相同的上下文. 如果是, 则第一级流水线不再从上下文状态存储电路中读取新的 Index 和 MPSS, 而是将 CX/D 延迟一个周期后直接送入下一级流水线; 第二级流水线也不再返回更新后的 Index 和 MPSS 到上下文状态存储电路

中, 而是根据这些数据直接从前一级流水线中读取概率估值参数进行计算. 即第一级和第二级流水线都忽略本次的上下文状态操作. 该技术既保证了流水线的正常进行, 也避免了逻辑冲突的发生.

如图 7 和图 8 所示, 本文和文献[14]都是采用三级流水线结构, 且第一级和第二级流水线基本功能和结构相似. 但本文采用了新的 RP 算法,

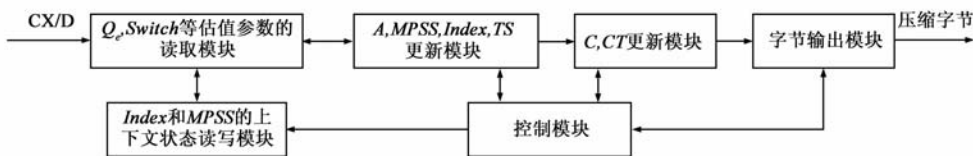


图5 整体结构

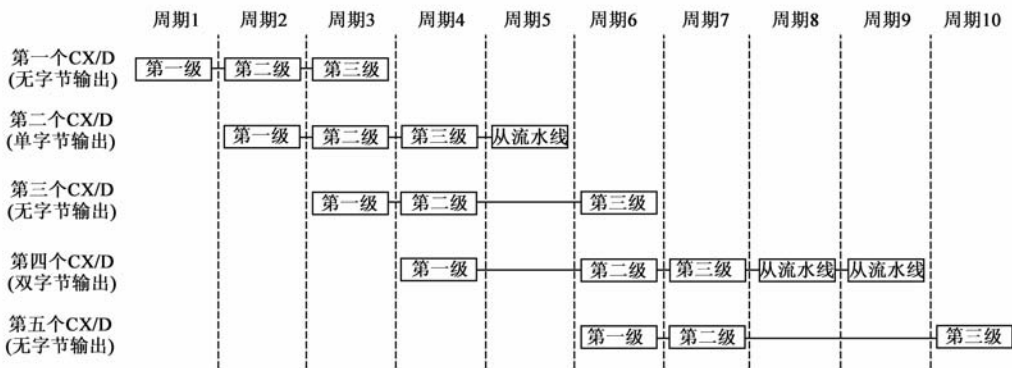


图6 流水线工作示意图

第三级流水线在完成  $C$  和  $CT$  的更新后, 仅需一个比较器来比较  $TS$  和  $CT$  即可判断出是否有字节输出, 判决逻辑简单; 通过从流水线专门输出编码字节, 实现了字节输出电路的复用. 而文献[14]在第三级流水线中包含了字节输出电路用于处理单字节输出, 并且并行了一个独立的检测和字节输出电路, 用于处理两编码字节输出的情况. 判决逻辑复杂, 没有复用字节输出电路. 因此相对于文献[14], 本文结构的电路面积更小, 且关键路径延时也更小. 这一点在下节将进一步讨论. 但考虑到算术编码时, 单字节输出占到了 10%. 因此相对于文献[14], 本文在相同时钟频率下 CX/D 的读入效率约低 10% 左右.

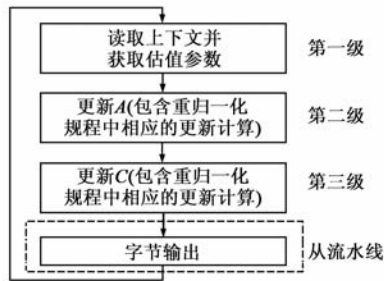


图7 本文流水线结构示意图

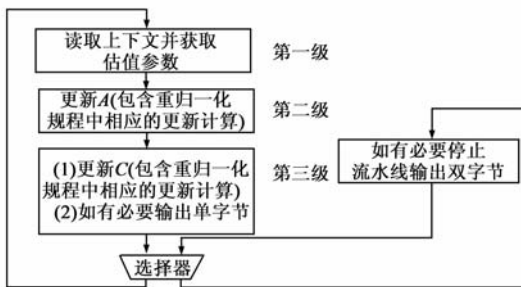


图8 文献[14]流水线结构示意图

### 3.2 关键路径分析

系统电路基于流水线结构实现, 因而整体性能取决于主、从流水线以及控制电路的关键路径. 控制电路基于状态机实现, 仅仅对各级流水线的控制信号和数据进行选择, 而不进行其他操作和计算, 因而关键路径最短. 从流水线对应的字节输出电路完成 BP 算法中的  $B$ ,  $C$  和  $CT$  更新, 共四个逻辑分支, 仅需一个选择器即可完成, 其关键路径也很短. 第一级流水线包括两个级联的查找表电路. 考虑到  $Q_e$  位宽较大, 因此求取  $Index$  和  $MPSS$  的查找表电路基于异步双端口 Ram 实现, 而求取  $Q_e$  和  $Switch$  等参数的查找表电路则基于同步双端口 Ram 实现, 有效缩短了关键路径. 系统的性能瓶颈集中在第二级和第三级流水线电路. 对于第二级流水线, 通过选择器并行完成  $TS$ ,  $A$ ,  $Index$  和  $MPSS$  的更新. 其中,  $A$  的位宽为 16 位, 且由于使用了新的 RP 算法, 需要在更新后执行移位计算, 因此  $A$  的更新计算成为本级电

路关键路径. 如图 9 所示,  $A$  和  $(A - Q_e)$  经过比较器比较, 输入选择器成为判决条件; 求得  $A\_temp$  后经移位计算, 得更新结果. 其路径长度包含一个减法器, 一个比较器, 一个选择器和一个移位器. 第三级流水线主要完成  $C$  和  $CT$  的更新. 如图 10 所示,  $C\_temp$  选择器的判决条件  $if\_stage2$  已在第二级完成, 直接读入即可. 而  $CT\_temp$  选择器的判决条件就是  $TS$  和  $CT$  本身. 判决条件简单, 与  $C\_temp$  的选择器并行进行. 由于  $C$  位宽为 28 位, 是本级位宽最大的数据. 因此  $C$  的更新计算成为本级的关键路径, 包含一个加法器, 一个选择器和一个移位器.

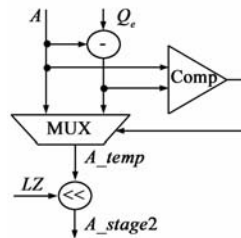


图9 第二级关键路径

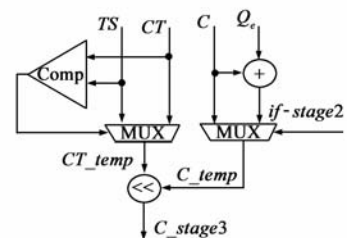


图10 第三级关键路径

文献[14]为了提高 CX/D 的读入速率, 在第三级流水线中包含了字节输出电路, 即将未进入重归一化规程前  $C$  的更新逻辑和 BP 算法中  $C$  的更新逻辑融合在一起, 从而延长了该级流水线的关键路径, 如图 11 所示, 显然劣于图 10 所示的关键路径.

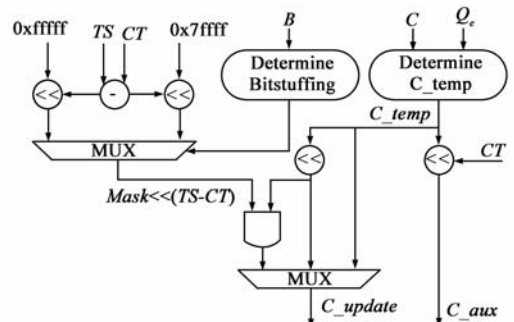


图11 文献[14]第三级流水线关键路径

## 4 实验结果

使用 Verilog 语言描述系统电路, 基于 TSMC 0.18 $\mu$ m 工艺, 使用 Synopsys Design Compiler 完成系统综合. 各级电路关键路径延时如表 2 所示. 选取国内外已发表文献中基于相同及相近工艺库的具有代表性的研究成果进行比较, 如表 3 所示.

表 2 各级电路关键路径延时

子模块	第一级	第二级	第三级	从流水线
关键路径(ns)	1.52	1.58	1.55	1.49

表 2 中, 关键路径延时集中在第二级和第三级流水线, 与理论分析结果一致. 表 3 中, 相对于文献[14], 本

文的系统时钟频率提高了 40%。但是,由于单字节和双字节输出的 CX/D 占总量的 10% 左右,因此实际增加的吞吐量约为 26%。同时,由于复用了字节输出电路,电路门数下降了 11%。

表 3 整体性能比较

结构	时钟频率 (MHz)	吞吐量 (Msymbols/s)	工艺库 ( $\mu\text{m}$ )	门数 (K)
本文	578	520	0.18	7.8
[11]	206	206	0.18	7
[12]	186	186	0.35	6.6
[13]	110	110	0.18	9.8
[14]	413	413	0.18	8.8

## 5 结论

本文在已有研究基础上,进一步研究了 JPEG2000 算术编码算法和电路实现。提出 RP 算法的一种新的顺序结构,并设计带有从流水线的三级流水线电路。基于 TSMC 0.18 $\mu\text{m}$  工艺的综合结果表明,系统时钟频率可达到 578MHz,吞吐量约为 520Msymbols/s,较国内外已发表文献的研究成果有了明显提高。可成为现有产品的替代方案以满足日益提高的高清数字图像应用需求。

## 参考文献

- [1] International Standard Organization. JPEG 2000 Part I Final Committee Draft Version 1.0, ISO/IEC JTCl/SC29/WGI [S].
- [2] 朱悦心,付昀,等.基于多级查询表的 JPEG2000 位平面扫描优化方法[J].电子学报,2004,32(5):810-813.  
Zhu Yuexin, Fu Yun, Wu Zongze, et al. A JPEG2000 optimized bit-plane scanning method based on multilevel query table[J]. Acta Electronica Sinica, 2004, 32(5): 810-813. (in Chinese)
- [3] 吴宗泽,郑南宁,等. JPEG2000 实时截断码率控制新算法及其 VLSI 结构设计[J].电子学报,2005,33(8):1457-1460.  
Wu Zongze, Zheng Nanning, Huang Yu, et al. New rate control algorithm truncating real time for JPEG2000 and its VLSI architecture design [J]. Acta Electronica Sinica, 2005, 33(8): 1457-1460. (in Chinese)
- [4] M Dyer, D Taubman, S Nooshabadi. Improved throughput arithmetic coder for JPEG2000[A]. Proceedings of IEEE International Conference on Image Processing [C]. USA: IEEE Press, 2004. 4: 2817-2820.
- [5] Y Z Zhang, C Xu, L B Chen. A dual-symbol coding arithmetic coder architecture design for high speed EBCOT coding engine in JPEG2000[A]. Proceedings of IEEE 6th International Conference on ASIC [C]. USA: IEEE Press, 2005. 1: 261-164.
- [6] M Dyer, D Taubman, S Nooshabadi, A K Gupta. Concurrency techniques for arithmetic coding in JPEG2000[J]. IEEE Trans-

actions on Circuits and Systems, 2006, 53(6): 1203-1213.

- [7] N Noikaew, O Chitsobhuk. Dual symbol processing for MQ arithmetic coder in JPEG2000 [A]. Proceedings of IEEE Congress on Image and Signal Processing [C]. USA: IEEE Press, 2008. 1: 521-524.
- [8] C Y Xiong, J H Hou, et al. Efficient fast algorithm for MQ arithmetic coder [A]. Proceedings of IEEE Conference on Multimedia and Expo [C]. USA: IEEE Press, 2007. 759-762.
- [9] B Min, S Yoon, J Ra, D S Park. Enhanced renormalization algorithm in MQ-coder of JPEG2000 [A]. Proceedings of International Symposium on Information Technology Convergence (C). USA: IEEE Press, 2007. 213-216.
- [10] C J Lian, K F Chen, H H Chen, L G Chen. Analysis and architecture design of block-coding engine for EBCOT in JPEG 2000 [J]. IEEE Transactions on Circuits Systems for Video Technology, 2003, 13(3): 219-230.
- [11] K Zhu, F Wang, et al. An efficient accelerating architecture for Tier-1 coding in JPEG2000 [A]. Proceedings of IEEE 7th International Conference on Solid-State Integrated Circuits Technology [C]. USA: IEEE Press, 2004. 3: 1653-1656.
- [12] Y Li, M Bayoumi. A three-level parallel high-speed low-power architecture for EBCOT of JPEG2000 [J]. IEEE Transactions on Circuits Systems for Video Technology, 2006, 16(9): 1153-1163.
- [13] Y Z Zhang, C Xu, W T Wang, L B Chen. Performance analysis and architecture design for parallel EBCOT encoder of JPEG2000 [J]. IEEE Transactions Circuits Systems for Video Technology, 2007, 17(10): 1336-1347.
- [14] M Rhu, I Park. Optimization of arithmetic coding for JPEG2000 [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2010, 20(3): 446-451.

## 作者简介



刘文松 男, 1983 年出生于江苏徐州, 东南大学信息学院电路与系统学科博士生; 2006 年和 2009 年分获中国矿业大学信息工程学士学位和电路与系统硕士学位, 2009 年至今就读于东南大学射频与光电集成电路研究所。研究方向为图像视频编解码芯片实现关键技术。

E-mail: xss4@163.com

朱恩 男, 1965 年出生于江苏南京, 东南大学信息学院电路与系统学科学科副主任, 教授, 博士生导师; 1992 年和 1995 年分获南京航空航天大学计算机专业硕士学位和自动控制专业博士学位, 先后就职于电子部 28 所、南京航空航天大学、中兴通讯, 2000 年至今就职于东南大学射频与光电集成电路研究所。参与的课题有: 国防大型电子工程、863 网络主题课题、航空基金项目、国防重点预研项目等。目前的研究兴趣包括: 电子系统设计理论, 光通信芯片实现关键技术, 图像/视频编解码芯片实现关键技术。

E-mail: zhenpro@seu.edu.cn