

文章编号 1004-924X(2009)09-2262-07

## 适于硬件实现的低复杂度图像压缩

徐 勇<sup>1,2</sup>, 徐智勇<sup>1</sup>, 张启衡<sup>1</sup>, 赵汝进<sup>1,2</sup>

(1. 中国科学院 光电技术研究所, 四川 成都 610209; 2. 中国科学院 研究生院, 北京 100039)

**摘要:**针对普通压缩算法和基于 DSP 或 PC 机的实现方法已不能满足高速和小体积图像压缩的要求,本文以硬件实现高速图像压缩为目的,提出一种基于 5/3 小波变换的低复杂度图像压缩算法。该算法首先采用 3 级二维小波变换去除图像相关冗余,并根据小波子带的变换增益对其进行最佳量化,然后对量化后的 LL 子带进行二维预测,最后针对小波系数概率分布的特点,采用自适应零游程编码联合指数哥伦布编码实现图像压缩。该方法在保证较高压缩质量的同时,具有低复杂度和硬件易实现的特点,通过 FPGA 最快可实现高达 175 Mpixel/s 的超高速图像压缩,为高速图像压缩器件的研制提供了算法基础。

**关键词:**图像压缩;小波变换;硬件实现;量化

**中图分类号:**TP391.4 **文献标识码:**A

## Low complexity image compression scheme for hardware implementation

XU Yong<sup>1,2</sup>, XU Zhi-yong<sup>1</sup>, ZHANG Qi-heng<sup>1</sup>, ZHAO Ru-jin<sup>1,2</sup>

(1. *Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu 610209, China;*  
2. *Graduate University of Chinese Academy of Sciences, Beijing 100039, China*)

**Abstract:** Because ordinary image compression algorithms and other compression methods based on Digital Signal Processors(DSPs) or PC machines can not satisfy the demands of high speed and small volume image compression applications, a low complexity image compression scheme based on the 5/3 Discrete Wavelet Transform (DWT) is proposed in this paper. Firstly, the correlation redundancy of image is excluded with 3-level z-dimension DWT. Secondly, the DWT coefficients are quantized according to the conversion gain of wavelet subband and the predict of LL subband is quantized. Finally, the image compression by combining adaptive zero run-length coding with Exp-Golomb coding is realized based on the probability distribution. Experimental results show the scheme has advantages in low computational complexity and hardware implementation, while it maintains relatively high performance in compressing various images. Moreover, it offers a high speed image compression of 175 Mpixel/s by FPGA, which provides an algorithm base for developing high speed image compression devices.

**Key words:** image compression; wavelet transform; hardware implementation; quantization

收稿日期:2008-08-18;修订日期:2008-10-23.

基金项目:国防创新基金资助项目(No. CXJJ-259)

## 1 引言

在工程应用中,常遇到对高分辨率、高帧频图像记录和传输的问题。如对于  $1k \times 1k \times 16 \text{ bit} \times 100 \text{ frame/s}$  图像记录,要求  $200 \text{ MByte/s}$  的接口速率,一般采用硬盘阵列的形式记录图像,记录设备难以小型化。虽然通过压缩能大幅减少数字图像的数据量以便存储和传输<sup>[1-2]</sup>,但目前压缩算法大都倾向于提高压缩性能而较少考虑算法复杂度和速度。另外市场上已出现多款图像压缩专用芯片,如 Tokyo Electron Device 公司的 Inrevium JPEG 芯片组 ( $63 \text{ Mpixel/s}$ ) 和 ADI 公司的 ADV202、ADV212JPEG2000 编码芯片 ( $65 \text{ Mpixel/s}$ ),但在高速应用中仍需要多片并行压缩才能满足速度要求,而且支持的视频格式和像素深度不能满足所有要求。因此,自主开发一种高速、高性能、能灵活配置的图像压缩器件是非常必要的。

为此,本文以硬件实现高速图像压缩为目标,提出一种低复杂度、高性能的图像压缩算法。该算法采用 3 级小波变换去除像素间冗余,并针对图像小波变换系数的概率和空间分布特征,设计了基于小波变换、低频预测、优化量化、自适应零游程编码(AZRL)和 Exp-Golomb 编码<sup>[3]</sup>级联的联合编码算法(简称 AZRL-EG 编码),较好地平衡了高压压缩性能和低复杂度之间的矛盾,为高速高性能的图像压缩器件的开发提供了算法基础。

## 2 图像小波变换

作为现代信号处理的重要手段,具有时频分析特性的小波变换在图像压缩、去噪、识别等领域获得广泛应用。在图像压缩领域,小波变换有明显的能量聚集作用,可有效去除图像的空间相关性。1992 年 Choen、Daubechies 和 Feauveau 提出了包含线性相位滤波器的 CDF5/3 小波和 CDF9/7 小波<sup>[4]</sup>,广泛应用于图像压缩,并被最新的图像压缩标准 JPEG2000 采纳<sup>[5]</sup>。在快速实现方面,Sweldens 于 1997 年提出了一种新的二代小波构造方法:提升方法(Lifting Scheme)<sup>[6]</sup>,计算量比基于滤波器组卷积的方法降低了一半,并

可实现“原位计算”,即在小波变换中无需附加内存,小波系数可直接覆盖原数据。正反变换的提升结构对称、实现简单、便于并行计算等优点使得提升结构成为实现离散小波变换的主流方法。

尽管在低码率时,CDF5/3 小波的图像压缩质量略低于 CDF9/7 小波,但由于实现 5/3 小波变换不需要乘法器,易于硬件实现,且可使算法支持无损压缩,故选用 CDF5/3 小波变换。基于提升结构的 CDF5/3 小波变换主要包含预测和更新两个操作,其方程如式(1)、(2)所示,式中  $x_{\text{extent}}$  表示原始信号  $x$  经过边界扩展后的序列, $\lfloor * \rfloor$  表示取整操作。本文中 CDF5/3 小波变换的边界扩展处理采用和 JPEG2000 相同的内嵌扩展模式,即偶数序号开始,奇数序号结束的模式<sup>[4,7]</sup>。

$$y(2n+1) = x_{\text{extent}}(2n+1) - \lfloor [x_{\text{extent}}(2n) + x_{\text{extent}}(2n+2)]/2 \rfloor, \quad (1)$$

$$y(2n) = x_{\text{extent}}(2n) + \lfloor [y(2n-1) + y(2n+1) + 2]/4 \rfloor. \quad (2)$$

## 3 小波系数的最佳量化

图像变换编码的通用框架由变换、量化、熵编码三部分组成<sup>[8]</sup>(如图 1 所示)。变换去除图像像素间的相关冗余,量化去除视觉冗余,而熵编码则去除概率冗余,一个变换编码的性能与这三部分密切相关。最佳编码可以在去除视觉冗余的同时使得失真最小化。

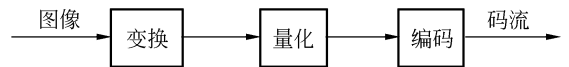


图 1 变换编码图像压缩系统的通用框架

Fig. 1 General frame work of transform coding

多级小波变换后的图像包含若干子带(如图 2 所示),子带变换增益  $G_b$  表示子带  $b$  的 IDWT 综合基矢量的平方范数<sup>[9]</sup>,它衡量了子带中每个系数失真对重建图像失真的影响程度,为最小化图像 MSE 失真,对变换增益较大的子带要采用较小的量化阶,考虑 Dead-Zone 均匀尺度量化,子带  $b$  的最优量化阶  $\Delta_b$  由(3)式给出<sup>[5]</sup>,式中的  $\Delta$  为基本量化步长,通过调整  $\Delta$  可控制码率和失真。对于正交小波变换,可以将变换增益近似为

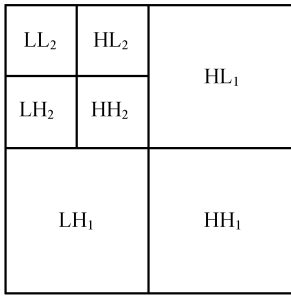


图 2 小波变换子带

Fig. 2 Sub-bands of DWT

(4)式,其中  $d_b$  表示子带  $b$  所属的变换级序号。

$$\Delta_b = \Delta \cdot \sqrt{\frac{1}{G_b}}, \quad (3)$$

$$G_b \approx 2^{2d_b}. \quad (4)$$

小波变换 LL 子带表示图像低频信息,HL、LH 表示竖向和横向边界信息,而 HH 子带表示了斜向的边界信息,一般来说,人类对 HH 子带的敏感程度低于其他子带,考虑到人类视觉加权和易于硬件实现,根据(4)式得到三级小波变换系数优化量化阶如表 1 所示。

表 1 三级小波变换子带优化量化阶

Tab. 1 Quantization step of 3 levels DWT subbands

子带/级数	1	2	3
LL	—	—	$\Delta$
HL	$4\Delta$	$2\Delta$	$\Delta$
LH	$4\Delta$	$2\Delta$	$\Delta$
HH	$8\Delta$	$4\Delta$	$2\Delta$

## 4 自适应零游程编码及 Exp-Golomb 编码

小波变换有效去除像素间的相关冗余,使能量聚集,因自然图像的分块平滑特性,故量化后小波系数在高频子带出现较多的全零区域,若单独编码系数样本,码率必然大于系数一阶熵<sup>[10]</sup>。为提高编码效率,有必要利用上下文相关信息编码以逼近其高阶熵。

根据高频子带存在大量全零区域的特点,采用自适应的零游程编码作为高阶扩展工具。本文

Samples: ... 1200012000000000010090000000029010 ...

Codes: ... 12001120081000900029010 ...

图 3 前向自适应零游程编码示例

Fig. 3 An instance of forward adaptive zeros run length coding

的自适应零游程包含两种模式:普通模式和零游程模式,编码时做前向自适应建模并在两种模式中转换,该规则比较简单:初始模式为普通模式,该模式下直接将输入样本输出,若连续输入 0 达到 2 个,进入零游程模式;在零游程模式无输出,只计数 0 个数,直到输入非 0 时退回普通模式,并输出 0 游程个数和该非 0 样本。图 3 给出部分样本零游程编码示例,下划线的数字表示处于零游程模式的样本及其零游程码,可以看出,如果信源中存在较多连续 0, AZRL 编码可以减少其样本数。

Exp-Golomb 编码<sup>[3,8]</sup>是一种可变长前缀码,其硬件实现简单,可用闭合公式解析码字,无需事先建立和存储码表,可通过比访问存储码字快得多的硬件计算产生码字,而且根据信源 PDF 函数灵活调整阶数  $k$  可以达到很高编码效率, Exp-Golomb 编码由于具有如上优点,已被用于 H. 264(AVC in MPEG-4)和我国自主提出的 AVS 标准中。其码字由两部分组成:一元码+组内标号。 $k=0$  时样本  $n$  的 Exp-Golomb 码的一元码长度由式(5)得到,组内标号码由式(6)得到,其中  $\text{bin}_m(n)$  表示  $n$  的  $m$  位二进制。式(5)、(6)码字连起来就可以导出硬件实现时编码公式<sup>[11]</sup>:若得到  $n+1$  的二进制首 1 位置  $j$ ,那么 Exp-Golomb 码字长度为  $2j-1$ ,由式(7)就得到整个 Exp-Golomb 码字  $C$ 。以上 Exp-Golomb 编码样本是非负的,采用式(8)将量化系数进行非负映射,将量化后小波系数  $c$  映射为非负样本  $n$ 。

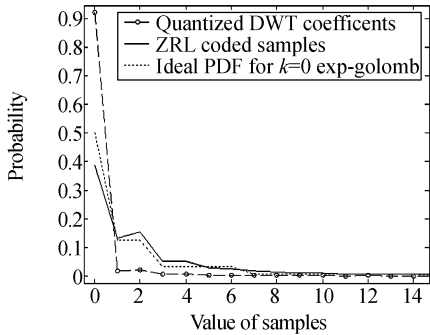
$$L = \lceil \log_2(n+1) \rceil + 1, \quad (5)$$

$$\text{IndexCode} = \text{bin}_{L-1}(n+1-2^{L-1}), \quad (6)$$

$$\text{Code} = \text{bin}_{2j-1}(n+1), \quad (7)$$

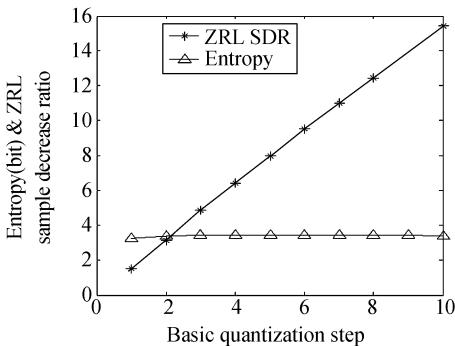
$$n = \begin{cases} 2c & c \geq 0 \\ -2c-1 & \text{others} \end{cases}. \quad (8)$$

实验表明,对于量化后的图像小波系数,零游程编码可以有效减少样本数,提高压缩比,图 4 展示了自适应零游程编码的性能,图 4(a) 包含 3 条曲线,分别是小波量化系数正映射后的概率密度函数(PDF)、零游程编码后的样本 PDF 和  $k=0$  的 Exp-Golomb 编码为最佳码时对应的理想 PDF。图 4(b) 为自适应零游程编码样本数量减少倍率和一阶熵随基本量化阶  $\Delta$  增大的变换规律。图 4(a) 表明:零游程编码后样本的 PDF 更符合  $k=0$  的 Exp-Golomb 编码的理想 PDF。图 4(b) 表明随着  $\Delta$  增加,ZRL 编码后样本数成倍减少,但其信息熵几乎不变。这样的话,采用 Exp-Golomb 编码对这些样本编码就可以快速高效地实现图像压缩。



(a) 在  $\Delta=4$  时 Lena. bmp 非负映射后系数和 ZRL 编码后样本的概率分布图

(a) PDF of lena after non-negative mapping and ZRL encoding with  $\Delta=4$



(b)  $512 \times 512$  书馆 Lena 图像随着  $\Delta$  增加,ZRL 编码后样本数量和熵变化规律

(b) Variation rule of sample number and entropy after ZRL coding

图 4 自适应游程编码的性能

Fig. 4 Performance of adaptive zeros run-length coding

### 5 低复杂度的图像压缩算法

根据以上分析即可得到一种低复杂度且易于硬件实现的图像压缩算法,图 5 为其流程框图。

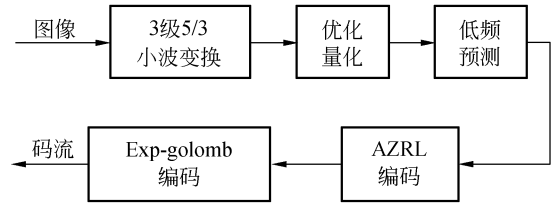


图 5 基于小波变换的低复杂度图像压缩算法框图  
Fig. 5 Block diagram of low complexity image compression algorithm based on wavelet transform

首先对输入的原始图像进行 3 级 5/3 小波变换;然后根据表 1 对各子带进行 Dead-Zone 均匀尺度量化得到量化系数;为进一步去除低频相关,对量化后的 LL 子带做二维一阶预测;此时在 HH,HL,LH 子带内会出现大量全 0 区域,以图 6 所示的扫描方式对量化系数进行自适应零游程(AZRL)编码,其输出样本再用  $k=0$  的 Exp-Golomb 编码进行熵编码。

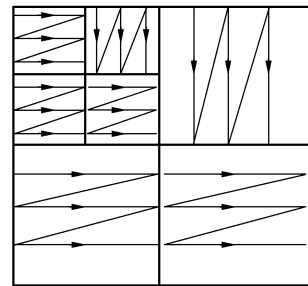


图 6 游程编码扫描模式

Fig. 6 Scan mode of run-length coding

自适应的 ZRL 编码是小波量化系数上下文相关模型建立的过程,并将其转换为与量化阶无关的概率分布,且符合  $k=0$  的 Exp-Golomb 编码的匹配概率,使得 Exp-Golomb 编码无需根据上下文调整参数,采用固定的  $k$  值就可以达到较高编码效率,这不仅有利于算法复杂度的降低,而且在硬件实现时可降低设计难度,提高工作频率及可靠性。

表 2 各图像在不同  $\Delta$  下  $k=0$  的 Exp-Golomb 编码效率Tab.2 Coding efficiencies of images in various  $\Delta$ 

图像/量化阶	$\Delta=1$	$\Delta=2$	$\Delta=4$	$\Delta=8$	$\Delta=16$
Lena. bmp	97.53%	97.70%	97.31%	96.84%	96.04%
Goldhill. bmp	96.62%	97.21%	97.51%	96.97%	95.56%
Zelda. bmp	97.37%	97.68%	96.75%	95.84%	94.25%

测试图像压缩实验表明,对于自适应零游程编码后的样本, $k=0$  的 Exp-Golomb 编码的一阶编码效率  $\eta$  达到 96% 左右(表 2)。一阶编码效率通过式(9)计算,其中  $H(X)$  是自适应零游程编码后样本的一阶熵, $\bar{l}$  是 Exp-Golomb 编码后的最终平均码长。

$$\eta = \frac{H(X)}{\bar{l}} \times 100\% . \quad (9)$$

由于熵编码的计算复杂度依赖于数据,难以进行估计,但在硬件实现时,可以通过编码所需的时钟数来估算算法的速度:JPEG2000 的 EBCOT 算法需要对小波系数约 10 个左右的位平面里的每个位平面进行 3 次扫描并用高复杂度的自适应算术编码进行编码,由于较低位平面的编码依赖于较高位平面编码,在硬件实现时难以实现并行设计,需要消耗大量的时间来进行多次扫描,如果图像尺寸为  $M \times N$  的话,执行小波变换需要大概  $M \times N$  个时钟,而执行编码约需要  $30 \times M \times N$  个时钟,另外复杂的控制逻辑也占用了大量的硬件资源。而本文在低频预测后仅需对变换后小波系数进行一次单像素扫描即可完成最佳量化、自适应零游程编码和 Exp-Golomb 编码的操作得到压缩码流,即完成熵编码仅需要  $M \times N$  个时钟,大大提高了编码速度。而且编码过程中无需计算和调整编码参数,这样就保证了系统的稳定性和高速实现的可能性,算法中涉及的小波变换、自适应零游程和 Exp-Golomb 编码,均有较成熟的硬件结构<sup>[11-14]</sup>,若采用流水线的硬件架构,编码速度可接近 1 Sample/Clk。从以上介绍可以看出,该算法具有结构固定、复杂度低、编码速度快的优点,通过 VLSI 或 FPGA 即可实现高速的图像压缩系统。

## 6 实验结果及分析

在 Visual C++6.0 中采用本文算法压缩几幅常用测试图像,并与最流行的 JPEG 标准及最

新的 JPEG2000 标准<sup>[5]</sup> 比较了压缩性能(图 7)。

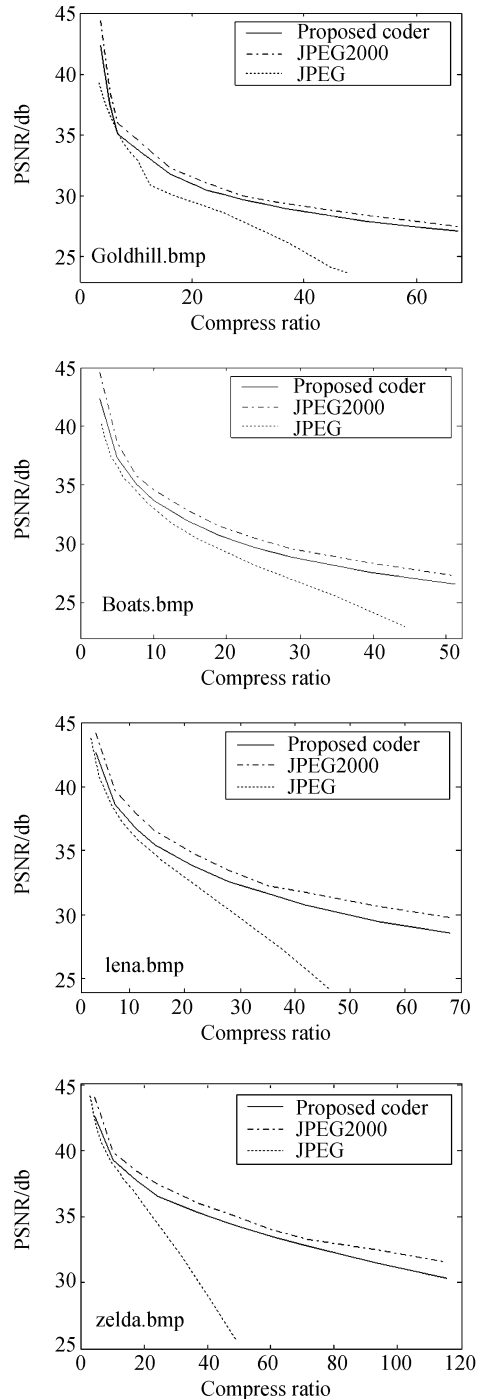


图 7 本文算法与 JPEG、JPEG2000 的图像质量比较  
Fig. 7 Comparison of our method JPEG2000 and JPEG

如图 7 PSNR-Compress Ratio 曲线所示,本文算法在压缩性能上优于 JPEG 标准,略低于 JPEG2000 标准。相比 JPEG2000 标准,本文为了获得较低复杂度且易于硬件实现,使得压缩性

能有较小下降,主要体现在以下两点:(1)本文基于统计意义给出最佳量化,与具体图像内容无关,而JPEG2000的PCRD-OPT算法<sup>[5]</sup>要根据具体图像编码数据得出码率优化。(2)JPEG2000采用高效但复杂的EBCOT编码算法<sup>[5,15]</sup>,熵编码效率优于本文的自适应ZRL-EG编码。可见,该算法较好地平衡了算法复杂度与压缩性能之间的矛盾,既能实现高速图像压缩,同时也具有较为满意的图像压缩质量。

基于该算法的高速图像压缩FPGA IP核已研制成功。该IP核采用脉动结构设计,包含控制器、图像分块器,3级小波变换器、预测器、联合编码器(集成了最佳量化、自适应零游程和Exp-Golomb编码)、码流打包器以及多个数据中转缓存,在主控制器的控制下,各处理器由数据中转缓存连接并进行脉动流水线操作,将图像分割为 $32 \times 32$ 或 $64 \times 64$ 的小块独立编码,最后打包成文件。选用Xilinx公司的xc5v1x50-3器件综合时频率可达到221.336 MHz,相当于175 Mpixel/s的编码速度,已远远高于前面提到的两款专用

芯片。

## 7 结 论

为了研制高速高性能的图像压缩器件,满足高速应用中对图像编码速度的要求,本文提出了一种复杂度低、易于硬件实现的图像压缩算法。该算法基于小波变换、低频子带预测、最小失真的优化量化以及自适应零游程编码结合 $k=0$ 的Exp-Golomb编码的熵编码,实现了高效快速的图像压缩。该算法具有以下特点:(1)复杂度低,编码过程仅扫描访问单像素;(2)没有复杂的上下文建模和编码参数调整,无需建立或存储码表,整个编码过程结构固定;(3)合理选择 $\Delta$ 可使整个编码过程不涉及乘除法和小数运算;(4)适应图像类型广,并有较好压缩性能。这些特点使得该算法非常适于硬件高速实现,基于FPGA实现时编码速度最高可达到175 Mpixel/s,为高速图像压缩器件的研制提供了算法基础。

## 参考文献:

- [1] 田宝凤,徐抒岩.一种适合星上应用的遥感图像有损压缩算法[J].光学精密工程,2006,14(4):725-730.  
TIAN B F, XU SH Y. A lossy compression algorithm of remote sensing image suited to spaceborne application [J]. *Opt. Precision Eng.*, 2006, 14(4):725-730. (in Chinese)
- [2] 赵洁红,孙文军.CCD遥感图像的小波分解特性及编码压缩[J].光学精密工程,2004,12(2):205-210.  
ZHAO J H, SUN W J. Wavelet transforms characteristics and compression coding of CCD remote sensing images [J]. *Opt. Precision Eng.*, 2004, 12(2):205-210. (in Chinese)
- [3] TEUHOLA J. A Compression method for clustered bit-vectors [J]. *Information Processing Letters*, 1978(7):308-311.
- [4] COHEN A, DAUBECHIES I, FEAUVEAU J C. Biorthogonal bases of compactly supported wavelets [J]. *Communications on Pure and Appl. Math*, 1992,45(5):485-560.
- [5] TAUBMAN D S, MARCELLIN M W. *JPEG2000 Image Compression Fundamentals, Standards and Practice* [M]. Kluwer Academic Publishers, 2001.
- [6] SWELDENS W. The lifting scheme: a construction of second generation wavelets [J]. *SIAM J. Math Anal.*, 1997,29(2):511-546.
- [7] TAN K C B, ARSLAN. Low power embedded extension algorithm for lifting-based discrete wavelet transform in JPEG2000 [J]. *C. Electronics Letters*, 2001,37(25): 1328-1330.
- [8] 吴乐南.数据压缩[M].2版.北京:电子工业出版社,2005.  
WU L N. *Data Compression* [M]. 2nd ed. Beijing: Publishing House of Electronics Industry, 2005. (in Chinese)
- [9] WOODS J W, NAVEEN T. A filter based bit allocation scheme for subband compression of HDTV [J]. *IEEE Trans. Image Process*, 1992(1): 436 - 440.
- [10] 姜丹.信息论与编码[M].合肥:中国科学技术大学出版社,2001.  
JIANG D. *Information Theory & Coding* [M]. Hefei: University of Science and Technology of China Press, 2001. (in Chinese)
- [11] WU D, GAO W, HU M Z, et al.. An Exp-Golomb

encoder and decoder architecture for JVT/AVS [C]. *In Proceedings of the 5th International Conference on ASIC*, 2003,2:910-913.

- [12] CHANG H C, CLICN L, CLIANG Y C, *et al.*. A VLSI architecture design of VLC encoder for high data rate video/image coding [C]. *Proc. of ISCAS, Orlando*, 1999:398-401.
- [13] LIAN C J, CHEN K F, CHEN H H, *et al.*. Lifting based discrete wavelet transform architecture for JPEG2000 [C]. *Proceedings of the 2001 IEEE International Symposium on Circuits and*

*Systems, ISCAS*, 2001,2:445-448.

- [14] DILLEN G, GEORIS B, LEGAT J D, *et al.*. Combined line-based architecture for the 5/3 and 9/7 wavelet transform of JPEG2000 [J]. *Transaction on Circuits and Systems for Video Technology*, 2003,13(9):944-950.
- [15] TAUBMAN D. High performance scalable image compression with EBCOT [J]. *IEEE Transactions on Image Proceedings*, 2000, 9(7): 1158-1170.

#### 作者简介:



徐 勇(1984—),男,云南红河人,博士研究生,主要研究领域为基于小波的高速图像数据压缩、基于 FPGA 的高性能信号处理、硬件计算结构设计等。E-mail:andrexu@163.com



张启衡(1950—),男,四川成都人,研究员,博士生导师,主要从事光电探测、目标跟踪、光电系统的研究工作。E-mail:qhzhang@ioe.ac.cn

#### 导师简介:



徐智勇(1967—),男,四川崇州人,研究员,硕士生导师,主要从事信号与信息处理、光电系统的研究工作。E-mail:xzy158@163.com



赵汝进(1982—),男,四川内江人,博士研究生,主要从事目标跟踪、姿态定位算法的研究。E-mail:zrj0515@163.com