

◎ 学术探讨 ◎

不等长信源的符号约束 MAP 联合译码算法

周丹丹, 邹 艳, 陆佩忠

ZHOU Dan-dan, ZOU Yan, LU Pei-zhong

复旦大学 计算机科学与工程系, 上海 200433

Department of Computer Science and Engineering, Fudan University, Shanghai 200433, China

E-mail: 052021139@fudan.edu.cn

ZHOU Dan-dan, ZOU Yan, LU Pei-zhong. Symbol-constrained joint source-channel MAP decoding of VLCs. *Computer Engineering and Applications*, 2008, 44(6): 39-41.

Abstract: Since Variable-Length Codes (VLCs) are inherently vulnerable to channel errors, traditional error-correcting codes are unable to solve the propagation-error problem efficiently. A new symbol-constrained joint source-channel decoding algorithm for VLCs is presented. By constructing double-tree VLCs sources, the propagation-error problem is evidently alleviated. Meanwhile, by applying symbol-constrained method to MAP algorithm, not only the a priori probability of VLCs sources is efficiently utilized, but also the Symbol Error Rate (SER) is reduced enormously. Compared with traditionally bit-constrained MAP algorithm, simulation results demonstrate that for JPEG source over AWGN channels, the gain in SNR of the method is around 0.8dB when SER reaches 1%.

Key words: joint source-channel decoding; variable-length codes; symbol-constrained; MAP decoding

摘 要: 由于不等长编码信源(VLCs)所固有的易于误码扩散的弱点, 传统的纠错编码并不能高效地解决其差错控制问题。提出了一种新的针对不等长编码信源的符号约束 MAP 联合译码算法, 利用 VLCs 双树信源的构造方法, 有效地抑制了 VLCs 的错误扩散; 同时, 通过将“符号约束”的思想应用于最大后验概率(MAP)译码算法, 不仅充分利用了信源先验信息, 而且极大地降低了误符号率。实验结果表明, 在误符号率为 1% 时, 对于在高斯加性白噪声(AWGN)信道传输情况下的 JPEG 信源, 提出的算法比传统的“比特约束”MAP 译码算法有将近 0.8 dB 的性能增益。

关键词: 联合译码; 不等长编码; 符号约束; MAP 译码

文章编号: 1002-8331(2008)06-0039-03 **文献标识码:** A **中图分类号:** TP39

1 引言

不等长信源编码(VLCs)如 Huffman 编码, 由于其极高的编码效率, 已被广泛应用于最近的图像和视频编码标准如 JPEG、MPEG-1/2/4 以及 H.264^[1]中。但在对不等长编码的信源进行译码时, 存在着错误传播和难同步问题。尽管一些传统的方法如自动重发(ARQ)可以降低信道错误的影响, 但这些方案导致了传输的延迟, 这在带宽受限的通信系统中造成了昂贵的通信代价^[2]。因此, 寻找另一种不等长编码信源的高性能译码方法, 已成为当今编译码领域中的热点课题。

为了提高不等长信源编码的译码鲁棒性, 文献[3, 4]中提出了一种软判决的 MAP 联合译码方法以降低误比特率。但这些算法在译码过程中产生了巨大的“格状图”, 导致复杂度过高, 无法被应用于较大的 VLCs 码表系统中。文献[5]提出了一种降低复杂度的查表的解决方法, 对 VLCs 的译码是在依次读入输

入比特的同时, 判断读入的若干比特是否恰好与某个码字匹配, 如果匹配就将对应的符号输出。由于 VLCs 的错误传播的特性, 这种“比特约束”的译码方法在误比特率已经降到很低的情况下, 误符号率仍然很高。另外, 由于文献[5]对 Markov 信源的依赖性, 造成了实用化困难。

本文提出的是一种基于双树构造的不等长编码信源的符号约束联合译码算法, 信道编码方式采用卷积码编码。针对 VLCs 信源的错误传播和同步问题, 本文对文献[5]提到的方法做出了两点改进: 通过在信源编码端构造 VLCs 双树源, 尽可能让比特错误引起的错误译码符号不出现在码表中, 有效地抑制了误码扩散。另外, 在译码端运用符号约束的联合译码算法(JSCD), 不仅充分利用了信源编码残留冗余, 而且尽可能地保证了符号和比特数目在发送和接收端的一致性, 极大地提高了译码性能。

基金项目: 国家自然科学基金(the National Natural Science Foundation of China under Grant No.90204013)。

作者简介: 周丹丹(1980-), 硕士生, 主要研究方向: 信息论与联合译码; 邹艳(1978-), 博士生, 主要研究方向: 纠错码与信息安全; 陆佩忠(1961-), 教授, 博士生导师, 主要研究方向: 纠错编码、信息论与通信技术、数学、信息安全等。

收稿日期: 2007-09-11 **修回日期:** 2007-12-10

本文下面的部分是这样组织的,在文章第2章介绍本文算法的传输模型;在第3章详细论述了 VLCs 双树信源的构造方法,以及针对 VLCs 双树信源的符号约束 MAP 联合译码算法;第4章给出实验结果和分析。

2 信源传输模型

传统的 VLCs 信源传输模型如图1所示。信源先进行 VLCs 信源编码,接着进行卷积码信道编码,经过 AWGN 信道,然后进行卷积码译码,最后进行 VLC 译码。

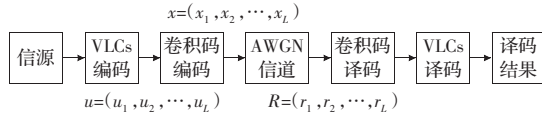


图1 传统的 VLCs 编码信源传输模型

本文的传输模型如下图2所示。为了使译码错误引起的误符号尽可能地不出现在码表中,先对待传输的 L 个信源符号进行 VLCs 双树编码,得到序列 $u=(u_1, u_2, \dots, u_L)$,其中 $u_i (1 \leq i \leq L)$ 表示原始信源符号 i 经过信源编码后的比特流。经过卷积码编码后的序列为 $x=(x_1, x_2, \dots, x_L)$,由于本文采用码率为 $1/2$ 的卷积码编码,所以 x_i 代表比特流对 (x_{i1}, x_{i2}) 。在经过 AWGN 信道后,得到接收序列 $R=(r_1, r_2, \dots, r_L)$,其中 r_i 代表接收的比特流对 (r_{i1}, r_{i2}) 。通过在译码端运用符号约束的 MAP 联合译码算法,不仅充分运用了信源先验信息,而且尽可能地保证了符号和比特数目在发送和接收端的一致性,从而提高了译码性能。

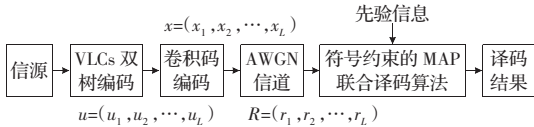


图2 基于双树构造的 VLCs 编码信源的符号约束 MAP 联合译码算法传输模型

3 VLCs 双树源的符号约束 MAP 联合译码算法

3.1 VLCs 双树信源编码

不等长编码如 Huffman 编码一般都是前缀码,由构造一棵二叉树来实现,任何一个码字都不会是另外一个码字的前缀。在译码过程中,单个比特的译码错误可能会引起译码路径的变化,而误码扩散的出现与否取决于误码是否会引起后续符号译码路径的连锁变化。也就是说,如果可以让前面出现的错误译码尽量地不影响后面的译码判决,就可以有效地抑制误码扩散。但是如果强行地对不等长编码的符号之间进行分割,会造成一定程度的冗余,降低传输效率。

对于通信系统中所构建的 VLCs 信源,由于信源符号的概率差异较大,构建的二叉树多为非对称二叉树,利用这个特性,本文提出了 VLCs 双树源的编码构造方法,即令奇数序号信源的符号码表和原始码表相同,偶数序号信源的符号码表为原始码表的取反,用于区分相邻的符号以抑制误码扩散。如图3所示。对于前面提到的信源符号集 $S=\{a, b, c, d, e\}$ 来说,对应的 VLC 码表为 $\{00, 01, 11, 100, 101\}$ 。那么奇数序号的 VLC 码表为 $S_o=\{00, 01, 11, 100, 101\}$,偶数序号的 VLC 码表为 $S_e=\{11, 10, 00, 011, 010\}$ 。可以看出,对于典型的非对称二叉树而言, VLCs 码表正表和反表的差异度明显增大,从而在保持信源压缩率不变的前提下扩充了码表,即使发生译码错误,该错误落在扩充

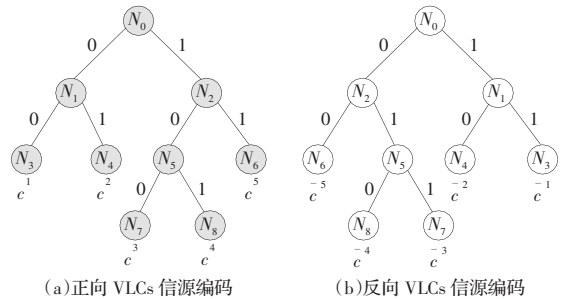


图3 VLCs 双树源

了的码表中的概率也会大大降低,从而有效地抑制误码扩散。

3.2 “符号约束”的 MAP 联合译码算法

3.2.1 “符号约束”的思想

为了保证符号和比特数目在发送和接收端的一致性,本文采用了“符号约束”的译码思想。图4给出了“符号约束”的示意译码方法。设给定一个简单的包含三个符号的符号集 $\{0, 10, 11\}$,横向表示符号的序号,方框里面的数字表示对应的比特序号。 $D(i, j)$ 表示比特序号为 j 的接收序列与有 i 个符号,长为 j 比特的所有码表组合之间的最小汉明距离。符号约束的 MAP 译码算法就是要在已知比特和符号数目的情况下,在图上寻找一条从起始位置到中止位置的一条接收序列和图中符号在相应位置之间的最短汉明距离的路径,然后回溯即得一条最优的符号路径。

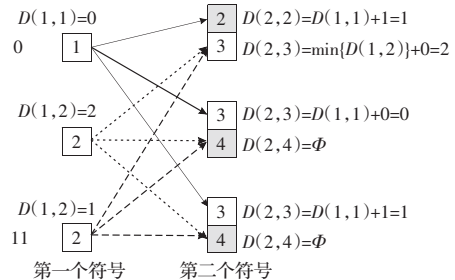


图4 符号约束的 VLCs 译码方法

假设传输的序列为 $\{0, 10\}$,包含三个比特二个符号。那么在译码时就可以直接剔除阴影方格所在的路径,直接选取 $D(2, 3)$ 的最小值,如图中斜体部分表示。然后再经过路径的回溯,即得加重箭头的译码路径,从而避免了由于译码符号数目的错误而引起的误码扩散问题。

3.2.2 符号约束的 MAP 联合译码算法

在对经过 VLCs 编码的信源进行译码时,假定信源符号的概率分布是已知的,利用这种先验信息可以提高译码性能^[9]。当使用卷积码进行信道编码时,一般采用基于 MAP 准则的 Viterbi 译码算法。

设信源符号集为 $\{c^1, c^2, \dots, c^C\}$, C 为待传输的符号集的容量。设 L 为要传输的一个分组内的符号数目,表示为 $Y=\{y_1, y_2, \dots, y_L\}$, $y_i=\{c^1, c^2, \dots, c^C\}$ 。状态序列 $s^{(i)}=(s_{i1}^{(i)}, s_{i2}^{(i)}, \dots, s_{iL}^{(i)})$, $s_{ij}^{(i)}$ 表示对于第 i 条可能的序列路径中第 j 个符号所处的状态。定义接收向量序列为 $R=(r_1, r_2, \dots, r_L)$ 。基于最大后验概率(MAP)准则,需在格图上寻找一条从 $s_{00}^{(i)}=0$ 出发,经过 L 个符号序列又回到全 0 状态的最佳路径,即要求出 $\max_i P(s^{(i)}|R)$,由于 R 与 i 无关

$$\begin{aligned} \max_i (s^{(i)} | R) &\Leftrightarrow \max_i (R | s^{(i)}) P(s^{(i)}) \Leftrightarrow \\ &\max_i (\ln \prod_{j=1}^L P(r_j | s_j^{(i)}, s_{j-1}^{(i)}) + \ln \prod_{j=1}^L P(s_j^{(i)} | s_{j-1}^{(i)})) \Leftrightarrow \\ &\max_i \sum_{j=1}^L (\ln(P(r_j | s_j^{(i)}, s_{j-1}^{(i)})) + \ln(P(s_j^{(i)} | s_{j-1}^{(i)}))) \quad (1) \end{aligned}$$

其中 $P(s_j^{(i)} | s_{j-1}^{(i)})$ 为先验信息, $P(r_j | s_j^{(i)}, s_{j-1}^{(i)})$ 为信道转移概率, 取决于传输信道和调制方式。在采用 BPSK 方式调制并在 AWGN 信道传输的条件下, 对于 1/2 卷积码来说, $P(r_j | s_j^{(i)}, s_{j-1}^{(i)}) = P(r_{j1} | x_{j1}) P(r_{j2} | x_{j2})$, x_{j1} 和 x_{j2} 是 t_j 时刻经过信源编码之后的两路输出向量: $x_{j1} = [x_{j1}^1, x_{j1}^2, \dots, x_{j1}^k]$, $x_{j1}^n \in \{0, 1\}$, $r_{j1} = [r_{j1}^1, r_{j1}^2, \dots, r_{j1}^k]$, $r_{j1}^n \in \{0, 1\}$, k 为第 j 个符号的向量长度, 且

$$P(r_{j1} | x_{j1}) = \prod_{n=1}^k \frac{1}{\sqrt{\pi N_0}} e^{-\frac{(r_{j1}^n - (2x_{j1}^n - 1)\sqrt{E_b/2})^2}{N_0}} \quad (2)$$

$P(r_{j2} | x_{j2})$ 的计算与上式类似, 其中 N_0 为噪声单边功率, E_b 为每信息比特的能量。

为了充分利用信道信息, 本文采用软判决译码, 对信道输出信号进行 $Q=8$ 级电平量化, 软判决译码器的任务就是寻找与接收序列 R 有最小软距离的路径。用 $d_s(r_t, x_t)$ 表示 t 时刻接收码字 r_t 与可能码字 x_t 的软距离, 式(1)可近似为求

$$M(s_{j-1}^{(i)}, s_j^{(i)}) = d_s(r_t, x_t) - w * \ln(P(s_j^{(i)} | s_{j-1}^{(i)})) \quad (3)$$

其中 $w = (\frac{E_b}{2N_0})^{-1}$ 为先验概率的权重, 卷积码译码器的任务就是要寻找累计分支度量值最小的路径:

$$\min_i \sum_{t=1}^L M(s_{t-1}^{(i)}, s_t^{(i)}) \quad (4)$$

对于传统的“比特约束”MAP 算法来说, 对 $\ln(P(s_j^{(i)} | s_{j-1}^{(i)}))$ 的计算需要根据先验符号概率计算得到先验比特概率; 而对于符号约束的联合译码算法来说, 对于每一个符号的先验概率的计算是基于符号内部比特概率的累加值的, 所以就没有了从先验符号概率计算比特概率的步骤, 即 $\ln(P(s_j^{(i)} | s_{j-1}^{(i)})) = \ln P(y_i)$ 。

4 实验结果

为了验证本文的算法对 JPEG 图像信源传输的效果, 实验采用文献[6]中的对 Lenna 图像进行统计过后的 8 符号的 Huffman 码表, 如表 1 所示。首先根据表中的统计概率随机生成一串序列构成信源, 然后进行卷积码编码, 通过图 2 的传输系统, 得出仿真结果。

表 1 JPEG DC 系数 Huffman 码表

DC 系数	DC 符号	统计概率 P	Huffman 编码
0	a	0.371 745	00
1	b	0.071 615	010
2	c	0.102 214	011
3	d	0.147 135	100
4	e	0.132 812	101
5	f	0.124 349	110
6	g	0.049 479	1110
7	h	0.000 651	11110

对于 VLCs 编码的信源符号, 比特错误出现的位置越靠

后, 误码扩散的范围也就越小, 因此, 减小分组长度可以降低误码扩散的影响; 而且对于符号约束的译码算法来说, 减小分组长度, 还可以减少回溯的次数, 降低算法的复杂度。但是如果分组过小, 卷积码编码器要频繁地清零, 导致复杂度较高, 因此, 在后面的实验中, 折衷地选取分组大小为 128。为了降低实验结果的随机性, 取 2 000 次随机实验的平均结果。

图 5 所示为 VLCs 信源在 AWGN 信道传输情况下的译码误比特率比较。可以看出, 本文提出的双树 VLCs 信源符号约束 MAP 联合译码算法的误比特率略高于传统的“比特约束”MAP 算法。这是因为传统的“比特约束”MAP 算法是按照比特来进行路径回溯的, 而本文提出的算法以符号为单位进行路径回溯, 减少了回溯的次数, 从而造成了误比特率的差异。但是对于不等长编码的应用而言, 评判译码性能的标准采用误符号率比误比特率更具有实际意义。因此本文主要采用误符号率来衡量整个系统的译码性能, 误比特率仅作为参考。

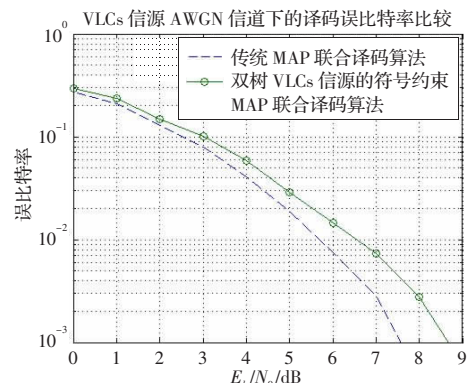


图 5 VLCs 信源 AWGN 信道下的译码误比特率比较

针对常见的生成多项式为 $[1+D+D^4, 1+D^2+D^3+D^4]$ 的 (2, 1, 4) 卷积码, 图 6 给出了 VLCs 信源在 AWGN 信道传输情况下 3 种译码算法性能比较。这 3 种算法分别为: (1) 传统的“比特约束”MAP 算法, 如文献[5]。在进行 VLCs 译码时, 是在依次读入输入比特的同时, 判断读入的若干比特是否恰好与某个码字匹配, 如果匹配就将对应的符号输出, 虽然误比特率可以很低, 如图 6 所示, 但是误符号率很高, 在 E_b/N_0 为 2 dB 时, 误符号率达到了 0.45 左右。(2) 符号约束的 MAP 算法。在译码过程中虽然确保了符号和比特数目传输过程的一致性, 但是由于码表的长度受限, 译码错误导致的误符号仍然很大概率地存在于码表中。(3) 符号约束的双树 MAP 联合译码算法。不但确保了符号和比特数目传输过程的一致性; 同时采用了 VLCs 双树源扩大了码

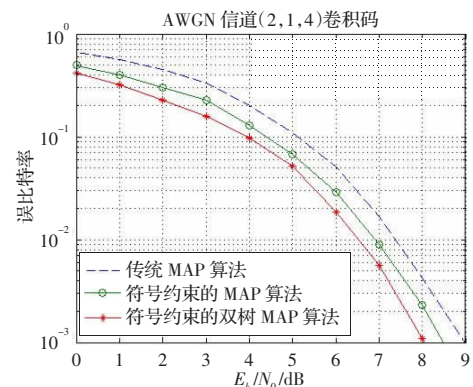


图 6 (2, 1, 4) 卷积码下几种算法的误符号率比较