

SAR原始数据压缩的快速矢量量化器

赵宇鹏^{1,2}, 万 锋^{1,2}, 雷 宏¹

(1. 中国科学院电子学研究所, 北京 100080; 2. 中国科学院研究生院, 北京 100039)

摘要: 回顾了矢量量化的各种快速算法, 指出其在应用到 SAR 原始数据压缩中的局限性。为了将矢量量化应用到 SAR 原始数据实时压缩, 提出一种基于查表的矢量量化器的设计方案, 其中整合了其他的矢量量化算法。最后用各种快速算法以及 BAQ 算法进行了对实际数据压缩的实验比较, 证明本算法在不降低量化速度的基础上, 提高了原始数据量化信噪比。

关键词: 矢量量化; 合成孔径雷达; 原始数据压缩; 快速搜索算法
中图分类号: O211.6 **文献标识码:** A

High speed vector quantizer in SAR data compression

ZHAO Yu-peng^{1,2}, WAN Feng^{1,2}, LEI Hong¹

(1. Institute of Electronics, Chinese Academy of Sciences, Beijing 100080, China;
2. Graduate School of the Chinese Academy of Sciences, Beijing 100039, China)

Abstract: Fast search algorithms are reviewed and their shortcomings in applications in the raw SAR data compression are addressed. For applying the vector quantization algorithm to raw SAR data compression, a vector quantization scheme is presented which is based on table look-up. The result of some experiments shows the efficiency of this algorithm.

Key words: vector quantization; synthetic aperture radar; raw data compression; fast search algorithm

1 引言

矢量量化作为一种有损的数据压缩算法, 由于其利用了原始数据的一些特点^[1], 如线性及非线性依赖、概率密度函数的形状、矢量的多维特性等。因此自 20 世纪 80 年代初提出以来, 就被证明比普通的标量量化更具优势。

矢量量化在 20 世纪 80 年代后半期被用在合成孔径雷达 (synthetic aperture radar, SAR) 的原始数据压缩上。可是在今天, SAR 数据压缩领域的 BAQ (block adaptive quantization) 还被广泛地应用着, 最关键的原因是矢量量化的速度无法应用在对传输速率要求非常高的场合。

提高矢量量化速度以及降低其复杂度的研究一直在进行着。这两方面的研究包括各种快速算法及各种约束矢量量化器。本文主要从快速搜索算法的角度出发, 利用数据间的相关性的快速搜索算法大多用于图像语音方面的压缩, 由于 SAR 数据的特殊性, 所以这里不予考虑。其它的快速搜索算法有: 部分距离法^[2,3]、投影法、超立方体法、最小最大法^[4]、子空间划分算法^[5]。快速算法的设计思路最初是减少乘法的次数, 但是无法缩减搜索的复杂度。子空间划分法仅用一步计算就可以排除 80% 左右的码字, 但是对剩余码字的搜索的复杂度也足以限制

矢量量化应用到 SAR 的数据压缩上。

当今一个实用的星载 SAR 系统^[6]的数据率大约为 200 Mbps, 平台上的固态存储器大约为 300 Gb。即使是实用的 BAQ 算法^[7]也已经被优化得无以复加。每一个下传的资料经过 2~3 次简单的计算, 就可以得到下传的码字。矢量量化即使是使用快速搜索算法也得进行很多步计算, 即使忽略搜索码书的过程, 其速度在高速的场合仍无法容忍。SAR 数据的高速率决定了一种量化方法必须依靠几次简单的处理才能够得到输入矢量的码矢。

2 算法

查表是一种减少计算步骤的好方法。在 BAQ 算法中就是利用查表达到快速求标准差的目的, 也是利用查表来直接找到输出码字^[7]。文献[8]中提到, 在 BVAQ 的过程中, 由于输入码字经过 BAQ 之后仅 4 位, 所以对于 4 维 4 位矢量量化, 所有可能的输入矢量的输出值存为一个表, 这种表仅需要 64 K 字节的存储量。在这个方法中, 矢量量化简化到仅仅需要一个简单的查表就可以获得输出码矢。如果把这种方法直接引入到矢量量化中, 就可以大大加快矢量量化的速度。

以同样的思路来设计一个矢量量化器, 计算所有可能输

收稿日期: 2003-05-11; 修回日期: 2004-04-23。

作者简介: 赵宇鹏 (1978-), 男, 硕士, 主要研究方向为雷达信号处理。

入矢量的码矢,通过查表输出。先来作一个简单的估算。给定 M 为比特采样数据,矢量维数为 K ,总码书的尺度为 N 。要存储的所有输入矢量的个数为 $2^{M \cdot K} \cdot \log_2 N$ 位。以常见的 8 比特采样 4 维矢量量化,码书含有 256 个码矢为例,存储量为 4 G 字节。这样庞大的存储量显然不能接受。实际的矢量量化每隔一段数据就要重新生成码书,所以必须先找出 $2^{M \cdot K}$ 这么多矢量的输出码字。这个数据段间隔一般为 100 M 左右。显然这种设计是极不合理的。

要想应用于实际,只有减少存储的输入矢量的个数。生成这个庞大的表时,其中的矢量并不是随机选取的,而是顺序增大每一维的值。在表中,位置相邻的矢量其输出矢量相同的可能性非常大。也就是它们很可能包含在同一个胞腔内。即使在不同的胞腔,也应该是相邻胞腔。用其中的一个代替另一个,不会带来很大的误差。从实际求得的全码表来看,至少连续 50 个左右的输出码矢是相同的。这样就为缩减全码表带来了方便。

在搜索全码表的过程中,并不进行完全的搜索,而是每隔一定的距离求一个输出值,这个距离如果设为 8,则全码表为 1 M 字节,如果设为 10,则全码表为 390 K 字节。这个数量级是现在的硬件水平完全可以接受的。传统的分裂法生成码书^[9]的速度在实际应用中无法满足要求,本文采用最大下降(MD)码书设计算法^[10,11],具体步骤如下。

步骤 1 将原始数据分块,采用 MD 法设计码书;

步骤 2 采用其它快速算法(本文使用子空间划分法)构造全码表;

步骤 3 查表输出。

从以上步骤可以看出,矢量量化已经简化为通过移位操作和简单的查表操作就可以找到输出码矢的程度了。与上述的各种快速算法相比,其计算量大大缩减。快速查表矢量量化方案如图 1 所示。

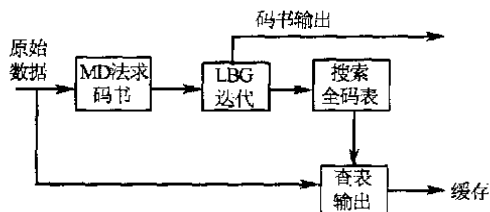


图 1 快速查表矢量量化方案

下面给出用各种快速搜索算法通过计算机模拟所得结果与 BAQ 量化原始数据结果的对比。其中 BAQ 算法的步骤如下。

步骤 1 原始数据绝对值累加;

步骤 2 直接查表输出。

算法流程如图 2 所示。

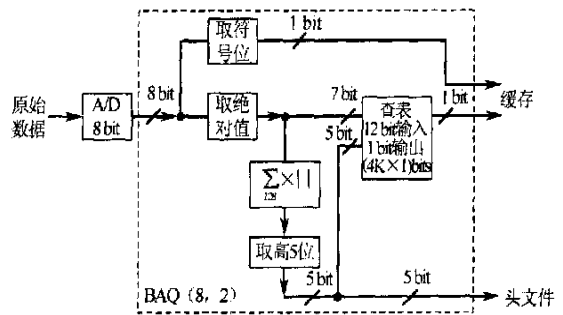


图 2 BAQ 算法流程

3 验证

选择 3 个分别为 80 M、160 M 和 320 M 的数据块,用计算机作压缩实验(其它矢量量化算法仅作了前两块数据的实验)。实验说明如下。

实验环境:硬件包括奔腾 IV 1.5G 处理器、256M RDRam、7 200 转 80 G 硬盘;软件环境为 Windows 2000 Pro、Visual Studio 6.0。

由于是计算机模拟,所以与实际的硬件实现有很多不同点。除了处理器的速度之外,硬件实现可以设计流水线结构,或者并行实现。

以减少乘法次数为出发点的算法在计算机上显示不出优势。

对子空间划分法作了改动,子码书直接由原码书冗余地生成,所以加快了速度,提高了量化信噪比,但是增大了存储量。

所有的矢量量化方法都使用部分距离法来提高搜索速度。

从表 1 量化信噪比看,所有的矢量量化方法都比 BAQ 高出一些,从量化的速度看,查表输出的矢量量化速度大致与 BAQ 的速度持平,远远高出其它的矢量量化方法,结合了速度和量化信噪比两方面的优势。而且后两块数据的量化实验表明,还可以减少额外的下传数据。抛开矢量量化的预处理部分(实际的矢量量化器两部分为并行工作)以及硬盘的读写时间(由于这种算法速度很快,硬盘的读写时间对实验的结果产生很大的影响),它的编码速度约为全搜索算法的 100 倍,而量化信噪比比全搜索算法下降得很少,比 BAQ 算法有较大提高。用较大的空间复杂性换取极高的时间复杂性取得了很好的效果。

表 1 各种压缩算法性能比较

量化方法	编码时间/s	量化信噪比/dB	额外下传数据/字节	内存占用(K 字节)
BAQ	9/18/33	11.167/8.857/7.154	1 000/2 000/4 000	32
全搜索算法	410/882	11.389/9.360/N/A	1 024	4
子空间法	133/241	11.319/9.356/N/A	1 024	24.3
投影法	260/540	11.116/9.068/N/A	1 024	12
快速查表法	19/23/31	11.333/9.337/8.079	1 024	1 000

4 结 论

从实验结果可以看出,本文提出的这种快速查表矢量量化的方法,把矢量量化的速度提高到了能够在星载 SAR 使用的水平,并且比 BAQ 量化提高了信噪比。其它的快速搜索算法同样在本算法中的搜索全码表部分起作用,比如本文使用的子空间划分法。这种算法明显的缺点是占用很大的存储量,在当前的硬件水平下,8 比特采样 4 维以上的矢量量化用硬件实现起来很难。

从矢量量化的实验过程来看,预计算的过程为 15 s,其中最慢的过程还是在 MD 算法给出初始码书之后,用 LBG 法迭代的这个过程。下一步的研究重点是缩减初始码书的寻找时间。

参考文献:

[1] Lookabaugh T D, Gray R M. High Resolution Quantization Theory and the Vector Quantizer Advantage[J]. IEEE Trans. on Information Theory, 1989, 35: 1020 - 1033.

[2] Bei C D, Gray R M. An Improvement of the Minimum Distortion Encoding Algorithm for Vector Quantization[J]. IEEE Trans. on Communications, 1985, 33: 1132 - 1133.

[3] McNames J. Rotated Partial Distance Search for Faster Vector Quantization Encoding[J]. IEEE Signal Processing Letters, 2000, 7(9): 244 - 246.

[4] Cheng De-yuan, Gersho A, Ramamurthi B, et al. Fast Search Algorithms for Vector Quantization and Pattern Matching[A]. Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing, 1984, 1: 1 - 9, 11.

[5] 李伟,林韵,胡征. 矢量量化的一种最佳快速搜索算法[A]. 第三届全国信号处理学术会议论文集[C], 815 - 817.

[6] Brů ě L, Baeggli H. RadarSat-2 Program Update[A]. Proc. of EUSAR '2002, 2002, 25 - 28.

[7] Curlander J C, McDonough R N. Synthetic Aperture Radar Systems and Signal Processing[M]. John Wiley & Sons, Inc, New York, 1991. 288 - 294.

[8] 杨云志,黄顺吉,王建国. 矢量量化在 SAR 原始数据压缩中的应用[J]. 系统工程与电子技术, 2002, 24(6).

[9] Linde Y, Buzo A, Gray R. An Algorithm for Vector Quantizer Design [J]. IEEE Trans. on Communications, 1980, 28: 84 - 95.

[10] Ma C K, Chan C K. Maximum Descent Method for Image Vector Quantisation[J]. Electronics Letters, 1991, 27(19): 1772 - 1773.

[11] 孙圣和,陆哲明. 矢量量化技术及应用[M]. 北京:科学出版社, 2002. 230 - 231.

(上接第 1505 页)

表 2 实验结果(II)

方法	超盒允许的 最大尺寸	警戒值	识别率/ %
GFMM	0.19	无	驱逐舰 83
			护卫舰 83
			导弹艇 98
	0.22	无	驱逐舰 83
			护卫舰 75
			导弹艇 94
无师 GFMM	0.19	0.88	驱逐舰 74
			护卫舰 68
			导弹艇 99
	0.204	0.905	驱逐舰 83
			护卫舰 79
			导弹艇 99

另外,无论是 ART2 的提出者 Grossberg 还是其应用者,所使用的警戒门限值 都相当高,文献[2]中 = 0.99,文献[3]中 = 0.992 ~ 0.995,而无师 GFMM 网络中参数 的值则通

常要小一些,一般在 0.9 左右。这充分体现了网络的稳定性和自适应性。

但在实验测试中也仍然存在需要进一步解决的问题。主要是网络的警戒值(门限值)、参数 和 应该怎样选取? 这些数值选得恰当与否对网络的实际性能有直接的影响。对于各种不同的样本集,为了使网络的性能达到最佳状态,最终所选取的网络警戒值和各参数值也有所不同。

目前所选取的网络警戒值和各参数值都是在具体的实验中测试出来的,其方法是对各种可能的数值进行穷举检验,直到选出最优值为止。用什么样的方法能够简单快捷地得到网络警戒值和其它参数值将是今后需要进一步探索的问题。

参考文献:

[1] Bogdan Gabrys, Andrzej Bargila. General Fuzzy Min-Max Neural Network for Clustering and Classification[J]. IEEE Trans. on Neural Networks, 2000, 11 (3): 769 - 783.

[2] Carpenter Gail A, Stephen Grossberg. ART2: Self-Organization of Stable Category Recognition Codes for Analog Input Patterns[J]. Applied Optics, 1987. 4919 - 4930.

[3] Chen S W. Neural-Fuzzy Classification for Segmentation of Remotely Sensed Images[J]. IEEE Trans. on SP, 1998, 45 (11): 2639 - 2654.