

快速可变速率 SAR 原始数据压缩算法

秦 蕾 王岩飞 胡晓新
(中国科学院电子学研究所 北京 100080)

摘 要 该文在原分块自适应矢量量化(BAVQ)算法基础上,提出了一种通过改变矢量码书大小及矢量维数来获得可变的编码数据率的改进措施。为了降低改进算法的计算复杂度,采用了数据块方差的查表法及矢量量化的快速搜索算法。对 SAR 原始数据进行压缩的结果表明,改进算法能够在不降低量化信噪比的情况下,获得更高的编码效率。不同数据率的解压缩数据生成图像,都不同程度地保留了原始图像中的细节信息。

关键词 SAR 原始数据压缩, 查表法, 快速搜索算法, 可变速率压缩

中图分类号: TN958, TN957.5

文献标识码: A

文章编号: 1009-5896(2006)02-0358-04

A Fast Variable-Rate SAR Raw Data Compression Algorithm

Qin Lei Wang Yan-fei Hu Xiao-xin
(Institute of Electronics, Chinese Academy of Sciences, Beijing 100080, China)

Abstract In this paper, based on the theory of the Block Adaptive Vector Quantization (BAVQ) algorithm, a variable-rate algorithm is proposed. In order to reduce the encoding computation complexity of the improved BAVQ algorithm, a fast VQ search algorithm and a look-up table to found quantizer control (threshold) are used. Using this improved algorithm to compress SAR raw data, a higher encoding efficiency is achieved and the SNR does not decrease in evidence. The decompressed images reserve most of details on the images resulted from raw data.

Key words SAR raw data compression, Look-up table, Fast search algorithm, Variable-rate compression

1 引言

分块自适应矢量量化(BAVQ)^[1]算法是一种有效的SAR原始数据压缩算法,它可以在数据压缩比较高的情况下,获得令人满意的数据压缩结果。但它只能进行单一数据率的压缩,不能满足通信系统对SAR图像质量的不同要求。也就是说,如果对SAR图像的质量要求较低,那么数据压缩系统只需将SAR原始数据压缩到较低的数据率即可,从而节省了系统的下行链路带宽;但如果通信系统对SAR图像的质量要求较高,那么较低的压缩数据率将不能满足对图像的高质量要求,因而需将SAR原始数据压缩到较高的数据率。为解决这种现实问题,本文提出了一种可变速率的改进BAVQ算法。在改进算法中增加了矢量码书的选择功能,按照对图像质量的不同要求,选择不同的矢量码书,从而实现对SAR原始数据的可变速率压缩。另外,由于BAVQ算法中结合了实现相对复杂的矢量量化(VQ)算法,使得其与实用的分块自适应量化(BAQ)^[2]算法相比,编码的复杂度较高,实现比较困难。为解决算法难于实现的问题,本文采用了数据块方差的查表法及VQ的快速搜索算法,不仅提高了编码效率,还增强了

算法的可实现性。本文首先分析了算法的改进原理,阐述了改进算法降低编码计算复杂度的方法,接下来给出用改进算法压缩SAR原始数据的结果,最后得出结论。

2 改进算法原理

以往的BAVQ算法只能对SAR原始数据进行单一数据率的压缩,不能满足通信系统对SAR图像质量的不同要求。为解决这个现实问题,本文提出了一种可变速率编码的改进算法。

用 N 代表码书的大小, K 代表矢量的维数,则VQ的编码数据率 R 可以表示为

$$R(\text{bit/采样}) = \frac{1}{K} \times \log_2 N \quad (1)$$

可见,编码的数据率只与矢量码书的大小和矢量的维数有关。因此本文的改进算法增加了一个矢量码书的选择功能,

并通过一个码速率选择信号控制码书选择电路,按照对SAR图像质量的不同要求,选择一个对应的码书,再进行VQ编

码,从而实现了可变速率编码的目的。

由于 BAVQ 算法中的 BAQ 量化器首先将原始数据进行了归一化处理,每块数据的统计特性大致相同,因此只需要非实时地生成一系列对应于不同数据率的,并与压缩系统无关的码书就可以。而且 BAQ 量化减小了数据的比特数,码书的尺寸都比较小,因此即使在星上增加多个码书,也不会增加星上存储设备的负担。

信源编码理论中的经典信息论指出,编码效率随着矢量维数的增大而增加。然而,编码的计算量 C 与码书大小和矢量维数的乘积成正比。由式(1)得

$$C = K \times 2^{R \cdot K} \quad (2)$$

其中 K 为矢量维数, R 为 VQ 的编码数据率。对于一个给定的数据率,编码的计算量随着矢量维数的增加以指数增长。因此,在设计码书时就需要兼顾编码计算复杂度的问题。码书的维数增大固然会提高编码的效率,但却以增加计算复杂度为代价,因此在本文的改进算法中,矢量的维数最大为 8 维。

本文主要考虑了数据压缩比大于等于 4 倍的情况,也就是说在采样数据 8bit 量化时,改进的压缩算法可以提供 2, 1.75, 1.50, 1.25, 1 bit/采样的 5 种压缩数据率。在本文的改进算法中,矢量维数最大仅为 8 维,最小为 4 维。长度最大的码书也只需 256 个矢量构成,最小则只由 32 个矢量构成。这样选择的主要原因是使压缩算法的计算复杂度降到最低,便于硬件实现。虽然当码书中仅有 32 个矢量的时候, VQ 量化引起的失真会增加,但是利用这样的小尺度码书进行 VQ 量化,编码搜索却非常简单,大大提高了 VQ 的速度,而且压缩结果表明, VQ 的性能不会因为采用了如此小的码书而受到大的影响。

由于 BAVQ 算法中结合了实现困难的 VQ,因此,为了降低改进算法的计算复杂度,以便于硬件实现,本文采用查表法和快速矢量量化编码方法来降低改进算法的计算复杂度。下面将详细介绍这两种方法的改进措施。

(1) 查表法 在 BAQ 量化器的实现中,由于直接计算数据块的方差是一项大运算量的工作,不仅增加了算法实现的复杂性,而且增加了硬件实现的难度,文献[2]中也提到了这一点。由于信号的平均幅度是分块内数据幅度的均值,因此相对于计算数据块方差的运算,平均信号幅度的运算简单得多。对于一个零均值的连续高斯分布数据来说,信号的平均幅度 $E|x|$ 与均方差 σ 有如下对应关系

$$\begin{aligned} E|x| &= \int_{-\infty}^{+\infty} |x| f(x) dx \\ &= 2 \int_0^{+\infty} \frac{x}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx = \sigma \sqrt{\frac{2}{\pi}} \end{aligned} \quad (3)$$

但实际处理的 SAR 原始数据是经过 A/D 量化的数据(一般量化为 8bit),因此 $E|x|$ 和 σ 的对应关系应改为

$$E|x| = 2 \sum_{n=0}^{L-1} \frac{x_n + 0.5}{\sqrt{2\pi}\sigma} \int_{x_n}^{x_{n+1}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \quad (4)$$

其中 L , x_n 分别为 A/D 量化器的阶数和转折点,对于本文处理的 8bit A/D 量化数据来说,式(4)可进一步化简为

$$E|x| = 127.5 - \sum_{n=0}^{127} \operatorname{erf}\left(\frac{n+1}{\sqrt{2}\sigma}\right) \quad (5)$$

其中 $\operatorname{erf}(\cdot)$ 为误差函数。

根据式(2)所示的平均幅度与均方差的对应关系,计算得到二者的对应曲线如图 1 所示。

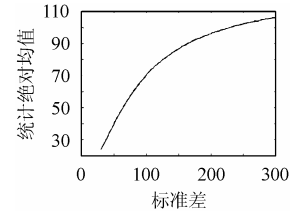


图1 平均幅度与均方差的对应关系曲线

按式(5)计算得到对应于不同平均幅度的方差值,并事先将这种对应关系存储在一个表中。对于某一个待量化的数据块,我们只需要首先计算数据块的平均信号幅度,查表得对应的方差,再进行数据量化就简单多了。引入方差查找表后,除了计算数据块的平均信号幅度这一项外,还需要一个查表的过程。数据块的平均幅度与方差对照表可按升序(或降序)的方式存放,这样可以大大节省查找时间。

采用这种快速查表的方法除了可以简化数据块方差的求解外,还有另外一个好处,就是如果在接收端也放置一个与发送端相同的对照表,那么信道中只需传输对照表中的数据块方差索引即可。这种方式类似于 VQ 中只传输码书中码字的索引,由于无需传输具体数值,增加了通信的可靠性。

(2) 快速矢量量化编码方法 VQ 以其固有的优越性一直是数据压缩算法考虑的对象,但其编码运算量很大,不利于 SAR 原始数据的星上实时压缩。信源编码理论中的经典信息论指出,编码效率随着矢量维数的增加而增加。但不幸的是编码的计算量随矢量的维数以指数增长。因此为了提高编码效率,减小计算量,本文改进算法中的 VQ 量化器采用了快速编码方法,可以在矢量维数较高的情况下,仅用较少的计算量就能实现矢量量化。

快速算法的设计思路最初来源于减少乘法的次数,但是无法缩减搜索的复杂度。而本文的这种快速搜索算法的思路

是不通过逐个比较而仅采用很简单的步骤,就可以排除大部分的候选码字,在剩余的很少的码字中搜索。为了实现这种快速搜索,就需要对码书的结构进行调整。预先将信号空间分成若干子空间,从而将搜索限制在一个相对较小的子码书中,这就是本文采用的子空间划分法^[3,4]。

子空间划分法的设计主要考虑两个问题,即子空间的划分和子码书的确定。子空间划分有两个基本原则:子空间的并集需填满整个信号空间;子空间的划分算法需简单。依据上述两个子空间划分的基本原则,本文采用了坐标划分法,即通过信号的符号矢量来确定子空间。对于 K 维空间,坐标划分法分出 $m = 2^K$ 个子空间。

首先用一个样本序列训练生成码书,再利用同一个训练序列得到子码书,具体的设计算法步骤描述如下:

(1) 给定训练序列 T (长为 L) 以及码书 B , 初始子码书:

$$B_i = \{C_j | C_j \in A_i, j = 1, 2, \dots, N\} \quad i = 1, 2, \dots, m$$

其中 C_j 为码矢, A_i 为子空间, 令 $n = 1$;

(2) 矢量 $X_n \in T$ 确定 A_j 使 $X_n \in A_j$, 由全搜索算法求出 X_n 的最近码矢 C_i :

$$d(X_n, C_i) = \min_{0 < j < N} d(X_n, C_j)$$

其中 $d(\cdot)$ 为失真测度, 一般选择 $d(X, Y) = \|X - Y\|_2$ 作为量失真测度;

(3) 如果 $C_i \in B_i$, 转到步骤(4)。否则扩充 B_i , 使得它包含 C_i ;

(4) 如果 $n < L$, 则 $n = n + 1$, 转到步骤(2);

(5) 如果 $B_i = \Phi$, 则 $B_i = B_{i-1} \cup B_{i+1}$ 。

算法的第(5)步是对空子码书的修正, 即采用邻近子码书填充的方法。这种空子码书修正的方法对低速率编码尤其重要。

本文采用了这种快速编码搜索算法后, 与全搜索算法相

比, 平均减少了 80% 的计算量。由此可见其高效性。

根据上述对改进算法原理的说明, 图 2 给出改进算法的编码原理框图。

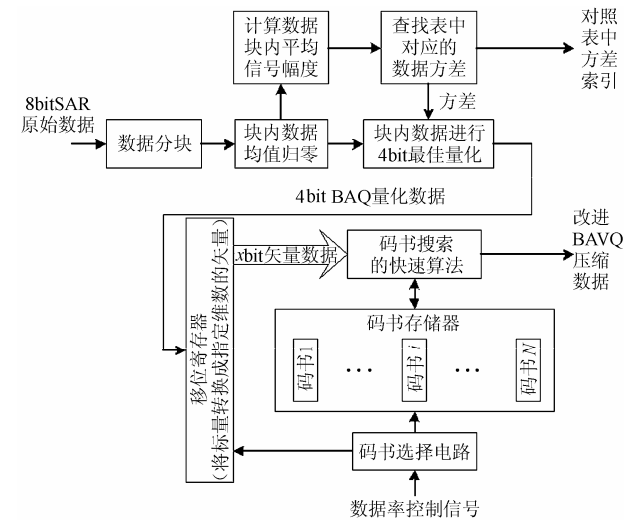


图 2 改进算法的编码原理框图

3 改进算法的编码结果

本文作了实际数据的压缩实验, 编码数据率分别为 2, 1.75, 1.50, 1.25 和 1 bit/采样, 下面给出分别采用全搜索矢量量化的 BAVQ 算法与本文的改进算法进行编码的运算时间, 如表 1 所示。

分析表中的计算结果可知, 随着码书尺寸的增大, 算法的执行时间增加, 说明算法的计算复杂度增加, 与前面分析的结论相符。与全搜索 BAVQ 算法相比, 本文改进算法的编码速度有很大幅度的提高。

下面给出利用本文的改进算法压缩 SAR 原始数据的结果, 如表 2 所示。

表 1 算法执行时间的对比表

编码数据率(比特/采样)	2	1.75	1.5	1.25	1
码书大小×矢量维数	256×4	128×4	64×4	32×4	256×8
全搜索 BAVQ 编码时间(s)	166	127	78	59	145
本文改进算法编码时间(s)	51	44	34	30	47

表 2 改进算法的实际数据压缩结果

压缩比 (倍)	数据率 (bit/采样)	SNR (dB)	PSNR (dB)	K	NMSE	ERMS
4	2	9.462	32.840	0.8722	0.1132	5.8376
4.57	1.75	8.325	31.704	0.8389	0.1471	6.6537
5.33	1.5	7.050	30.428	0.7838	0.1972	7.706
6.4	1.25	5.839	29.214	0.7092	0.2609	8.8621
8	1	5.016	28.394	0.657	0.315	9.74

表中 SNR 为量化信噪比, PSNR 为峰值信噪比, K 为逼真度, NMSE 为归一化均方误差, ERMS 为平均误差。压缩结果表明, 在压缩比很大(如 8 倍)时, 本文的改进算法仍然可以获得很高的量化信噪比值。

下面给出对应于不同数据率的成像结果, 如图 3 所示。

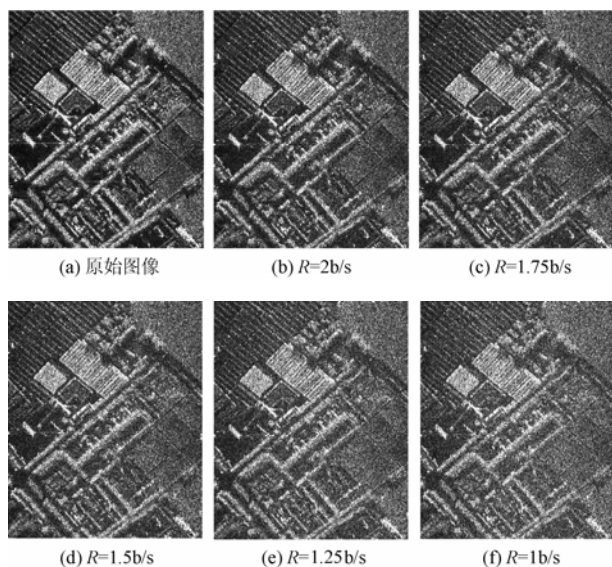


图3 解压缩数据成像结果及其与原始图像的比较

这是一幅细节比较丰富的地域图像。由解压缩数据成像结果表明, 数据压缩比逐渐增大时, 图像质量几乎没有多大变化, 都能不同程度地保留图像中的细节信息。由此可见本文改进算法的高效性。

4 结束语

本文通过在 BAVQ 算法中增加矢量码书的选择功能, 对 SAR 原始数据进行不同数据率的压缩, 从而改进了原 BAVQ 算法。为了降低改进算法的计算复杂度, 增强硬件可实现性,

采用了方差的查表法, 从而提高了 BAVQ 算法中 BAQ 量化部分的编码效率, 在 VQ 量化部分采用了快速编码的算法, 与全搜索编码算法相比, 在降低编码计算复杂度的同时, 几乎不降低数据压缩的质量。利用本文改进的 BAVQ 算法压缩实际 SAR 原始数据的结果表明, 算法的执行时间比较短, 在数据压缩比很高的情况下, 仍然可以获得较高的量化信噪比, 而且解压缩数据生成图像能够不同程度地保留图像中的细节信息。由此可见, 本文的改进算法是一种高效的 SAR 原始数据压缩算法。

参考文献

- [1] Moreira A, Blaeser F. Fusion of block adaptive and vector quantization for efficient SAR data compression. Proc. IGARSS, Tokyo, 1993: 1583 – 1585.
- [2] Kwok R, Johnson W T K. Block adaptive quantization of Magellan SAR data. *IEEE Trans. on Geoscience and Remote Sensing*, 1989, 27(4): 375 – 383.
- [3] Cheng R Y, *et al.*. Fast search algorithms for vector quantization and pattern matching. Proc. ICASSP, California, 1984: 9111 – 9114.
- [4] 李伟, 林韵, 胡征. 矢量量化的一种最佳快速搜索算法. 第三届全国信号处理学术会议论文集, 1988: 815 – 817.

秦 蕾: 女, 1979 年生, 硕士生, 研究方向为合成孔径雷达数据压缩算法。

王岩飞: 男, 1963 年生, 博士生导师, 研究方向为微波成像雷达系统及其理论、数字信号处理等。

胡晓新: 女, 1977 年生, 博士生, 研究方向为合成孔径雷达数据压缩算法。