

部分饱和 SAR 原始数据压缩¹

赵宇鹏*** 万锋*** 雷宏*

*(中国科学院电子学研究所 北京 100080)

** (中国科学院研究生院 北京 100039)

摘要: 分块自适应量化 (BAQ) 算法在实际应用中碰到的问题是 SAR 原始数据含有较多饱和成分的量化。文中分析了 BAQ 对这类数据产生较大误差的原因。根据数据的特点, 提出了一种对传统 BAQ 进行改良的压缩算法——部分饱和 BAQ(FSBAQ)。实验表明, 这种方法在不改变压缩比的情况下, 对存在一定饱和的数据都能提高 4.5dB 以上的量化信噪比。文中给出了一条选择量化形式的曲线, 并给出用传统 BAQ 算法及该文给出的算法对一块 SAR 原始数据进行量化和结果对比。

关键词: 合成孔径雷达, 原始数据压缩, 饱和, 分块自适应量化

中图分类号: TP751, TN951 **文献标识码:** A **文章编号:** 1009-5896(2004)03-0489-06

Compression on Fractional Saturation SAR RAW Data

Zhao Yu-peng*** Wan Feng*** Lei Hong*

*(Institute of Electronics, Chinese Academy of Sciences, Beijing 100080, China)

** (Graduate School of the Chinese Academy of Sciences, Beijing 100039, China)

Abstract A quantization method is presented to aim at the fractional saturation Synthetic Aperture Radar (SAR) raw data. Norm distribution is the basis of Block Adaptive Quantization (BAQ) algorithm. But in the application of BAQ, some trouble may encounter, one of these is fractional saturation of raw data. A compression algorithm based on the traditional BAQ algorithm is presented, which can select different ways to quantify the raw data. This algorithm reduces the error and increases the Signal to Quantization Noise Ratio (SQNR) without influence the compression ratio. Finally, a block of real SAR raw data are processed by the two algorithms and the result is compared with each other.

Key words SAR, Raw data compression, Saturation, Block Adaptive Quantization (BAQ)

1 引言

SAR 在工作过程中会产生大量的原始数据, 在向地面发送的过程中, 由于传输速率的限制, 必须在下传之前, 对其进行处理, 以期减少数据率。

SAR 原始数据的压缩算法研究虽然已经有很长时间, 种类很多, 效果也各不相同, 但是由于受到实现的复杂度以及算法的速度等问题的制约, 当前实际应用的压缩算法仍然很少。BAQ^[1] (Block Adaptive Quantization) 算法因其简单快速而被广泛采用。虽然 BAQ 算法应用广泛, 但是在实际的使用中也会遇到各种问题。在对原始数据进行采集量化的时候, 某些情况下, 例如雷达飞经城市上空, 原始数据会产生过多饱和, 原始数据不再符合正态分布, BAQ 就失去了理论基础, 用传统的方法对其量化, 势必会产生很大的误差。因此有必要具体研究数据的特点, 对传统的算法进行改良。

2 问题提出

BAQ 算法的基础, 认为 SAR 原始信号的数据是在距离向和方位向具有缓变方差零均值的

¹ 2002-10-10 收到, 2003-03-10 改回

高斯分布信号^[2]。将原始数据分成若干小块,估算这个小块内的方差。应用这个方差将块内数据归一化,使之符合均值为零,方差为1的标准高斯分布。以(0,1)高斯分布为基础,预先计算好量化电平以及输出电平。最后拿归一化的数据与量化电平比较得到量化后的码字。在地面,根据接收到的数据、各数据块的方差以及输出电平恢复出原始数据。图1为BAQ算法的流程图。

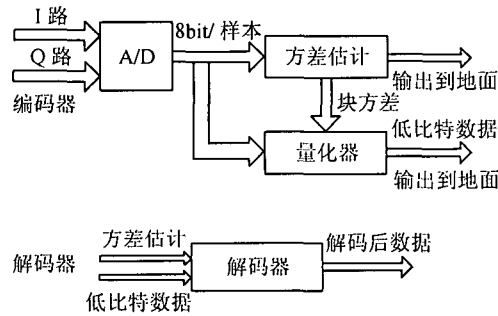


图1 BAQ流程图

从以上对BAQ算法的分析过程中可以看出,原始数据的分布近似服从高斯分布是个非常重要的条件。实际的大多数数据也很好符合这个条件。但是例外的情况也很多,其中比较常见的就是数据中含有较多的饱和。图2和图3是一块SAR原始数据的波形图和数据分布的统计直方图。从图中可以看出,这块数据含有较多的饱和成分,这是由于某些特定的情况,比如由于A/D转换器的阈值调节不当,造成A/D饱和,很多采样数据等于A/D变换器的最大值和最小值,这个时候,上述的BAQ算法的基本条件已经不能满足,如果还是用传统的BAQ进行量化的话,会造成很大的误差。

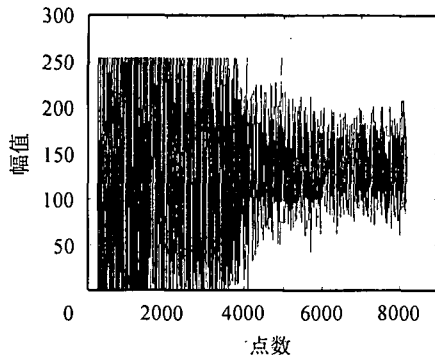


图2 不规则原始数据波形图

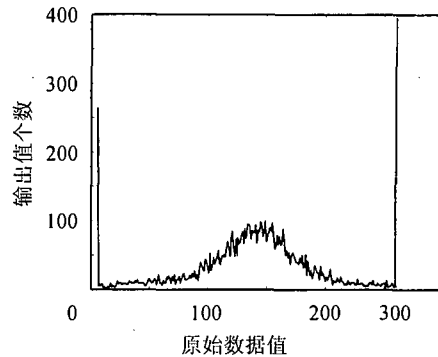


图3 不规则原始数据统计直方图

下面具体分析产生很大误差的原因。

BAQ量化器的量化噪声功率为^[3]

$$E[\varepsilon^2] = E[(s - y)^2] = \sum_{k=1}^M \int_{s_k}^{s_{k+1}} (s - y_k)^2 p(s) ds \quad (1)$$

在量化中使用的计算数据块的标准差的公式^[1]:

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

其中 $\operatorname{erf}(x)$ 是误差函数: $\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-x^2) dx$ 。

式(2)将 A/D 转换器输出数据的幅度统计均值与高斯分布的标准差联系起来^[1]。在实际的压缩、解压缩过程中一般先将这个映射制成表,以便在压缩和解压缩时采取查表的方式,以达到快速处理的目的。

当数据块的标准差为 50,饱和成分很少的情况下,进行 3bit 量化,由式(1)可以得出 BAQ 的理论量化误差为 15.2dB。而当饱和的输出值 (± 127.5) 概率增加到 0.2 的时候,仍然按照上述的 BAQ 方法进行计算,标准差为 74,且理论量化误差下降为 12.9dB。

下面生成一块指定方差为 σ^2 ,均值为 0 的高斯分布的数据,并且取整,使之与 SAR 的原始数据具有同样的特性。表 1 是标准差计算结果对比。表中饱和数据占总数据的百分比是根据实际的原始数据的饱和数据百分比给出的,为 20%。表 2 给出标准差为 50 的数据加入不同百分比的饱和之后通过 BAQ 量化计算所得的标准差。

表 1 数据加饱和和前后量化结果比较

未加饱和数据标准差	未加饱和的量化误差	加饱和的量化误差	噪声增加百分比 (%)	加饱和后标准差
30	5.5306	7.9590	43.87	56.5171
40	7.2547	10.5521	45.25	65.1807
50	8.6996	16.0215	84.16	74.4831
60	9.9543	12.6879	28.65	86.1449

表 2 不同百分比标准差计算结果

饱和百分比 (%)	4	10	16	22
计算标准差	54.4726	61.9067	69.5489	78.0330

在对添加了饱和的数据的量化过程中,根据式(2)计算标准差。由于式(2)是根据原始数据近似符合高斯分布为前提推导出来的,由于前提条件已经不能满足,如果按照这个方法推算原始数据的方差肯定存在问题。从表 1 和表 2 可以看出,加入这些饱和的数据,根据查表方式得到的方差变大,导致量化区间展宽,结果造成量化误差增大。从以上的分析可以看出,有必要对饱和过多的数据进行特别处理,使之不影响非饱和数据方差的计算。

3 改良算法的提出

这里的分析中,原始数据是 8 位 A/D 转换器量化,最大值为 255,最小值为 0。从原始数据的分布直方图上可以看出,数据的最大值和最小值的出现频率远远高于其他中间值的出现频率,甚至几倍于数据的均值的出现频率。在除了最大值和最小值的区域,中间区域基本上符合高斯分布。BAQ 的设计思想就是在数据分布密集的区域,取较小的量化区间,在数据分布稀疏的区域,取较大量化区间。这样在均方误差意义下,全局量化误差达到最小。根据这个思想,由于最大值和最小值是局部数据非常集中的区域,所以有必要用专门码字表示。对于中间值,从统计直方图上看,数据仍然能够很好地符合高斯分布,所以仍然使用传统的 BAQ 算法对其量化。从原始数据的波形图可以看出,数据含有过多饱和只在一定的区域,并不会延伸到所有的数据块。所以要对进行量化的数据先进行判断,判断饱和数据是否多到需要进行特别的处理。

以下是具体实现步骤:接收到原始数据之后,先对其进行分块,然后统计块内饱和数据在块内所占的百分比,同时统计出非饱和数据的方差。预先设置一个百分比,当饱和数据所占百分比小于此值时,对数据块进行传统的 BAQ 量化,当饱和数据所占百分比大于此值时,采用对

饱和数据单独编码的方法，其余数据仍然使用根据高斯分布计算得出的门限电平进行量化。下传的数据包括量化后原始数据、各数据块的方差和各数据块的量化方法指示。在地面，根据量化的方法和数据块的方差恢复出原始的数据。

表 3 是 3bit 常规的 BAQ 量化分别在正常情况和含有饱和数据的情况下，输出电平所代表的原始数据的幅值。从表中可以看出，含有饱和的数据所代表的幅度值已经超出了原始数据的幅值范围，所以表示超出范围的码字就不会用到。3bit 量化的方法应用于含有饱和的数据其实已经浪费了两个码字，所以本量化方法利用两个码字来专门表示饱和的数据绝对不会增大量化误差。反而原来会产生很大量化误差的饱和数据现在是无误差的量化。

表 3 量化电平、输出电平及其代表的幅值

	最佳量化电平	最佳输出电平	代表幅值 (无饱和)	代表幅值 (含饱和)
1	0.000	0.2451	0~30.036	0~40.048
2	0.5006	0.7560	30.036~63	40.048~84
3	1.050	1.344	63~104.88	84~139
4	1.748	2.152	104.88~∞	139~∞

这里比较常规的 BAQ，FSBAQ 带来的好处有以下两点。

(1) 在计算方差的时候，除去饱和的数据，使量化的方差更好地符合非饱和数据，减少非饱和数据量化的误差；

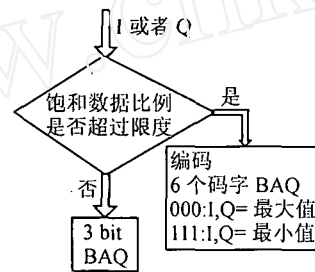


图 4 饱和 BAQ 算法流程

(2) 饱和数据很大的量化误差变成了无误差。

以下是这种算法的流程图 (以 3bit 量化为例)：

本算法的实质是在实际应用这种算法的时候，根据数据块的统计特性，自适应地选择均方误差最小的量化方法。当数据比较符合高斯分布的时候，使用常规的 3 bit BAQ，当含有较多的饱和数据时，使用 6 个码字的 BAQ 算法对非饱和数据进行量化，其他的两个码字用来专门表示两个饱和值。下传的数据包含了量化后的原始数据、数据块的方差和量化方法指示。比常规 BAQ 算法每一个数据块多了一个指示字符，几乎相当于没有增加数据量。如果数据完全没有饱和情况，那么本算法也完全不会影响量化的质量，所以可以说是一种很通用的量化方法。

下面仅从算法层面估计两种算法的计算量。从算法的流程来看，FSBAQ 的算法由于增加了判断步骤，所以计算量有轻微的增加。以一个数据量为 M 的数据块进行 3bit 量化为例，传统的 BAQ 需要进行 M 次累加，一次查表确定标准差，3 次乘法计算量化门限，最后通过 $M \sim 3M$ 次比较得出当前数据的输出值。假设数据饱和和百分比为 δ ，FSBAQ 需要 M 次比较， $(1-\delta)M$ 次累加，两次查表，两次乘法计算量化门限， $M(1-\delta) \sim 2M(1-\delta)$ 次比较得到输出值。可以看出，新的算法对于存在一定饱和， M 在 4000~250000 的范围内，计算量增加的并不是很大。在最后的比较输出阶段，计算量甚至有可能减少。所以总体上，计算量增加不多。

在上面算法的流程中, 第一步就是要检查数据统计特性, 判断饱和和百分比适用于哪一种量化形式。但是在用给定方差的数据进行实验的过程中发现, 传统的 BAQ 算法的量化误差随着饱和和百分比的增长并不是单调的, 而是有一个升高、降低再升高的过程。而饱和 BAQ 算法随着饱和和百分比的增长是单调递减的, 如图 5 所示。

从图中可以看出, 普通 BAQ 可能在两个区间的量化误差可能低于饱和 BAQ。所以, 并不能通过简单的计算来判断, 而是需要预先给出一条标准差-饱和和百分比的判断曲线, 通过计算数据块的非饱和数据的方差, 以及统计出饱和数据的百分比, 与预先给出的这条曲线对照, 就可以判断使用哪种量化形式。随着饱和和百分比的增大, 两种量化方法的量化误差会两次相交。但是在两次相交的区间中, 两种量化方法的误差相差很小, 所以在本文给出的判断曲线, 就取第一个两条曲线的交点为临界点。判断曲线如图 6 所示。

在实际使用中, 把图 6 所示的曲线存为标准差-饱和和百分比表。从原始数据得到没有达到饱和和数据的标准差, 查表得出转换百分比, 与实际数据饱和和百分比进行比较, 就可以选择量化的方式。此时根据 (1) 式, 用 FSBAQ 进行量化饱和占到 20% 的数据, 理论信噪比可达到 16.9dB, 比 BAQ 理论信噪比提高了 4dB。

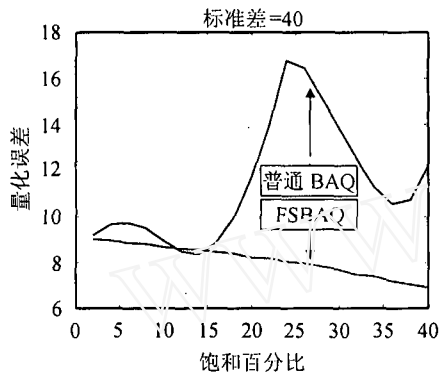


图 5 两种量化方法的量化误差曲线

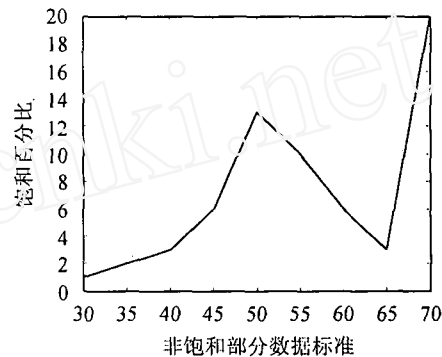


图 6 判断曲线

4 实际数据检测效果

为了验证这种算法效果, 本文以实际的 SAR 原始数据作了两组实验, 分别是: 3bit BAQ 量化, 3bit 饱和 BAQ 量化。表 4 是对这两种量化方法结果的比较。从表中可以看出, 对于这一块特定的数据, 改良的 BAQ 算法能提高量化信噪比 5.33dB。其它的数据块分析结果表明, 这种算法对于存在 10% 以上饱和的数据都能提高 4.5dB 以上的量化信噪比。

表 4 两种方法对原始数据量化的比较

量化方法	BAQ	FSBAQ
σ	91.220	91.221
MSE	24.207	13.111
SQNR	11.523	16.849

传统的 BAQ 量化方法是根据均方最小误差准则得出的, 而 FSBAQ 是在数据含有饱和和成分时继续贯彻最小均方误差准则, 衡量的标准就是 SNR, 所以这里仅仅给出了 SNR 的对比。

从原始数据的统计特性来看, 当饱和达到 30% 以上时, 非饱和部分的数据呈现近似三角分布。但是饱和过多并不是数据没有用处, 仍要进行量化下传。在处理非饱和数据时, 两种算法效果接近。但是由于 FSBAQ 对饱和部分的数据作了特殊的处理, 而传统的 BAQ 在处理饱和和数据时会造成很大的误差, 所以总体上 FSBAQ 优于 BAQ。

实验的结果表明, 这种改良的 BAQ 算法对于降低含有较多饱和的原始数据的量化误差有很大的帮助。而从这种算法的简单性可以看出它是一种现阶段比较实用的 SAR 原始数据压缩方法, 可以很方便的应用到机载及星载 SAR 上面。

5 结束语

本文对 BAQ 算法应用于部分饱和的 SAR 原始数据量化进行了研究。用一块实际的 SAR 数据, 验证了这种 FSBAQ 的效果。从本文给出的对实际数据的量化结果对比表明: 本文提出的方法在压缩此类部分饱和的 SAR 原始数据要优于传统的 BAQ 算法。由于 FSBAQ 糅合了有损压缩和无损压缩的方法, 所以经过此方法量化后的数据, 不能再进行有损处理, 比如矢量量化等。另外这种方法只对 3bit 及以上的量化方法比较有效。

参 考 文 献

- [1] Kwok R, Johnson W. Block adaptive quantization of Magellan SAR data. *IEEE Trans. on Geosci. and Remote Sensing*, 1989, 27(4): 375-383.
- [2] Curlander J C, Mcdonough R N. *Synthetic Aperture Radar Systems and Signal Processing*. New York, John Willey & Sons, Inc., 1991: 288-294.
- [3] 刘永坦等. 雷达成像技术. 哈尔滨: 哈尔滨工业大学出版社, 1999: 206-213.

赵宇鹏: 男, 1978 年生, 硕士生, 主要研究领域为雷达信号处理。

万 锋: 男, 1979 年生, 硕士生, 主要研究领域为雷达信号处理。

雷 宏: 男, 1963 年生, 研究员, 主要研究领域包括微波成像、天线技术、无线通讯。

www.cnki.net