

大规模油藏三维显示系统关键技术的研究

狄翠萍, 申闫春, 刘 双

(北京信息科技大学 计算机学院, 北京 100192)

摘要:本着提高三维模型实时渲染速度和用户视觉效果的目的,在不增加硬件资源消耗而单从提高已知数据信息利用率和利用速度的基础上,提出了利用临时表提高数据查询速度,降低分辨率以提高显示速度,利用双向链表存储和边缘算法提高视觉效果等大规模油田数据处理的新方法。这些方法在大规模油田数据模型的三维显示中得到了积极的应用。

关键词:临时表;低分辨率;双向链表;模拟退火

中图分类号:TP391.9

文献标识码:A

文章编号:1673-629X(2010)02-0237-04

Study of Key Technologies of Large - Scale 3 - D Display System

DI Cui-ping, SHEN Yan-chun, LIU Shuang

(School of Computer, Beijing Information Science & Technology University, Beijing 100192, China)

Abstract: In order to improve the speed of real-time rendering of three-dimensional model and users' visual effects, bring forward a number of new ways of data processing about large-scale oilfields, which include the using of temporary table, lower-resolution display to enhance speed and the storage of a two-way linked list and the edge algorithm to improve the visual effects of large-scale oilfields, not increasing hardware resource consumption, but improving the utilization rate and utilization speed of known data from the information. Those ways are applied positively in the display of large-scale three-dimensional data model.

Key words: temporary table; low-resolution; two-way linked list; simulated annealing

0 引言

油藏模拟的关键问题就是油藏数据的管理、调度与实时渲染,从而最大限度地提高油藏模型的实时显示速度和最大程度地提高视觉效果。大规模数据的处理一直是油藏描述与油藏模拟领域研究的一个重点,也是一个难点^[1]。虽然众多文献提出了许多油藏模型网格粗化以及提高内存性能的理论与方法,但都效果不佳。将一个油藏数据较准确地以三维世界中模型的方式展示给用户,需要提高已知数据的信息利用率和利用速度^[2]。

目前,在数据处理方面,最常见的一种现象就是,当数据量很小时,系统可以正常运行,一旦数据量增大到一定程度,系统就会出现异常。针对这一问题,通常采用的解决方法是更换处理能力更强的计算机来满足用户要求^[3]。在油田等一些大部门,甚至采用高速度

和高性能的巨型机来满足大规模数据处理的实际要求。这样,随着数据量的不断增加,需要不断地更新硬件来满足要求,而硬件的升级会给用户带来很大的经济负担,所以从改善硬件方面考虑提高大规模数据的处理速度已经没有实际意义。

1 临时表技术在油藏模型中的应用

数据库的一个显著功能就是可以将本地表检索分割,这使得大型表和索引可以得到高效的管理,将大规模的油藏数据分割成小数据分别放到数据表里,这样就变成对小数据的处理^[4,5]。但是,借助数据库处理大数据的能力也不是没有限度的,对大数据量的实时处理问题仍然是大型数据库应用系统中的共性问题。油藏数据属于一个大数据,对一个存有大量油田数据的表进行频繁的实时性要求很高的并发操作时,系统在某一时间段内无法完成用户的响应,导致数据库死锁或拥塞,最终导致系统停止服务也再所难免,采用临时表来缩减同时需要处理的数据量能够从很大程度上解决这个问题。

临时表是一种比较特殊的表结构,临时表不能永久地保存数据,只是一个能够很快读取数据表中数据

收稿日期:2009-05-31;修回日期:2009-08-14

基金项目:北京市教育委员会科技计划项目(KM200811232006);科技型创新基金项目(BT2008-13)

作者简介:狄翠萍(1982-),女,硕士研究生,研究方向为虚拟现实技术、图形图像处理;申闫春,博士后,从事计算机网络与应用、流媒体技术、虚拟现实技术方面的研究。

集的媒介。虽然将数据集插入到临时表中时,会给系统带来额外的开销,但是,当临时表

中的数据集远小于源数据表中的数据时,在进行数据连接操作或对数据集进行频繁读操作会比直接对源数据表操作时系统的性能提高几倍甚至几十倍。

油藏数据是地下油、气、水等多种属性的集合。由于用户一般要查看的不是全部属性的模型,所以采用的方法是将油、气、水等属性分别写入到临时表中,然后再对临时表进行连接运算(见图 1)。

当用户要求查看某一属性模型时,直接调用相应的临时表,而不必直接调用原始的油藏数据模型这个大数据表,只有当用户要求要看全部油藏属性模型时才直接调用原始数据模型。油藏模型临时表中的数据集远小于原始的数据表中的数据,而且原始数据表中还存在着很少被查询的一些属性,因此,借助临时表,可以大大提高对大数据表访问性能,从而提高实时处理的性能。

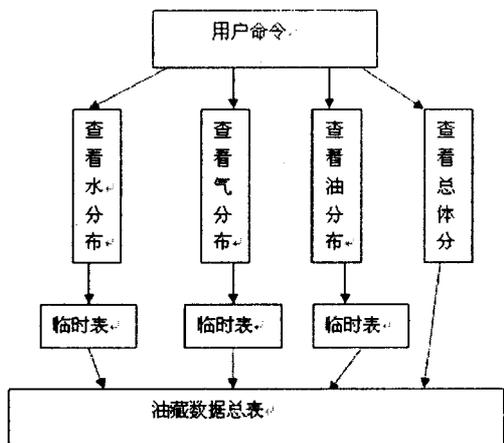


图 1 临时表调用关系举例

2 数据粗化节省内存资源

尽管将油藏大规模数据分类抽取并采用临时表技术进行处理能够有效地提高对大数据量的实时处理性能,但这种分析处理减少的只是需要同时处理的数据量,并没有减少存储空间。进行数据量大数据分析抽取分组时,会对系统中的实时数据请求产生很大的冲击,严重时也会出现请求超时或死锁现象。

要解决上面提出的问题,还必须通过一定的算法剔除一些不需要的数据,减少最终需要处理的数据量。在百万接点数据处理模型中,可以将实际的油藏数据粗化,数据粗化的本质特征是求均值或者近似值。文中采取的措施就是将油田数据取整的近似算法。可

662	1:	C GRID BLOCK: I = 35, J = 1, K = 1				
663	1:	82422.25	97778.00	1307.46	82422.17	97732.00 1304.89
664	1:	82401.57	97733.50	1304.99	82401.82	97779.38 1307.59
665	1:	82425.08	97779.70	1309.84	82425.08	97733.49 1306.75
666	1:	82402.24	97735.11	1307.16	82402.31	97781.20 1310.12
667	1:	C GRID BLOCK: I = 36, J = 1, K = 1				
668	1:	82422.17	97732.00	1304.89	82422.08	97686.00 1303.26
669	1:	82401.31	97687.61	1303.05	82401.57	97733.50 1304.99
670	1:	82425.08	97733.49	1306.75	82425.08	97687.29 1304.93
671	1:	82402.17	97689.03	1304.55	82402.24	97735.11 1307.16
672	1:	C GRID BLOCK: I = 37, J = 1, K = 1				
673	1:	82422.08	97686.00	1303.26	82422.00	97639.94 1303.43
674	1:	82401.05	97641.67	1303.82	82401.31	97687.61 1303.05
675	1:	82425.08	97687.29	1304.93	82425.08	97641.08 1305.95
676	1:	82402.09	97642.94	1306.26	82402.17	97689.03 1304.55

图 2 油藏数据表截图

数据粗化后的油藏模型应当是健壮的和有抵抗力的,对于这样的数据抹去小数点后的数字几乎是没有影响的。而且由浮点型转换成整型不但可以降低计算难度,而且还可以大大减小存储空间。同时,采用 OpenGL 为作图工具,如果要看到模型,需要定义一个可视空间,可视空间以外的物体才能看到。这么庞大的油藏模型如果要想映射到屏幕上,那么可视空间要够大,而且,如果需要看到模型的全部,视点位置要离模型足够远,这么一来,浮点型和整型的差别是完全可以忽略的。

另外,在比较靠近外层的原油属性表(见图 3)中可以看出,由于水、油、气等都逐渐减少,所以许多属性值都为 0,这时,对于值为 0 的属性,完全可以只做判断而不做处理,即某个属性为 0,直接以某种颜色填充,不需要存储数据。这样一来,又大大节省了资源。

```

00003840h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ;
00003850h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ;
00003860h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ;
00003870h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ;
00003880h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ;
00003890h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ;
000038a0h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ;
000038b0h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ;
000038c0h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ;
000038d0h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ;
000038e0h: 41 D7 39 5E 42 78 9F 96 42 86 97 C 42 41 17 74 ;
000038f0h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ;
00003900h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ;
00003910h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ;
00003920h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ;
00003930h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ;
00003940h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ;
00003950h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ;
00003960h: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 ;
  
```

图 3 油藏属性的二进制表示

3 降低分辨率和采用颜色索引提高渲染速度

虽然高分辨率的模型能真实地反映地下油藏的分布,但是往往受计算机运行速度和存储容量的限制,高分辨率在大数据模型上的应用非常困难[6]。而油藏模型是为了将地下采集到的油藏数据直观地反映给用户,只要能分辨出地下油藏中原油储量及其油、气、水等属性的分布即可,所以强调高分辨率没有多大意义。

此项目中的油藏数据只有几十个属性,所以经过分辨率重建,用低的分辨率来渲染模型,大大节省了 CPU 的处理时间。

OpenGL 支持 RGBA 和颜色索引这两种颜色模式。两种颜色模式,计算机都必须为每一个像素保存一些数据。不同的是,RGBA 模式中,数据直接就代表了颜色;而颜色索引模式中,数据代表的是一个索引,要得到真正的颜色,还必须去查索引表,虽然搜索索引表占用一定的时间,但是如果索引表中的数据是很少的,那么搜索需要的时间可以忽略,而且索引颜色的主要优势是占用空间小,每个像素不必单独保存自己的颜色,只用很少的二进制位就可以代表其颜色在颜色表中的位置,花费系统资源少,图形运算速度快。

在索引颜色模式中,OpenGL 需要一个颜色表。利用这个颜色表可以调出很多种颜色,调出的颜色不会多于颜色表的中颜色的数量。颜色表的大小是很有限的,一般在 256~4096 之间。在使用索引颜色方式进行绘图时,总是先设置颜色表,然后选择颜色。在油藏模型中要表示的油藏属性有限,颜色表中保存的颜色就代表各油藏属性。

4 双向链表边缘渲染提高视觉效果

边缘算法是一种优先显示边缘轮廓的算法,经常用在一些远程显示系统中提高视觉效果。文中的数据是精细油藏模型,获取边缘的方法是将模拟退火算法融入到遗传算法中,由于对原始细网格油藏模型做数值模拟往往需要大量的 CPU 机时,成本太高,甚至是不可能实现的,故可以利用油藏模型中采样得出的部分数据逐步逼近真实值,模拟出整个油藏模型的边缘轮廓。利用模拟退火算法得出解空间,利用遗传算法对解空间进行优化,避免了遗传算法容易提前收敛的缺陷以及模拟退火法搜索较盲目的缺点。

模拟退火法是模拟冷却晶体的过程,模拟退火算法是一种求解优化问题的通用算法,它将可行解及其目标函数值视为物质体系的状态和能量,通过模拟物质体系的退火过程求优化问题的最优解。另外,遗传算法除了具有强的鲁棒性和并行性之外,也具有全局优化的特点。将模拟退火算法和遗传算法结合起来应用在数据优化问题上,会得到更好的效果。

4.1 模拟退火算法

模拟退火算法是从乱数中寻找临近的点^[7],若找到的点比当前点好,更逼近真实值,则取之,否则依照概率判定是否取之。

算法流程如下(见图 4):先设定一些参数,接着随机产生一个初始的目前解,并计算它的目标函数值

$f(x)$ 。以目前解为中心对解空间做随机扰动,产生一个扰动解,其目标函数值为 $f(x')$ 。若接受,则以该扰动解取代目前解作为该次叠代的解。

由于我们最终需要的是油藏的边缘轮廓,所以,比较靠近边缘的目前解是有效的。但是,最终的边缘解的获得是由油藏内部的点通过模拟得来的,而油藏内部和油藏边缘的地质中各属性是有差别的。目前解的寻找要尽量避开油藏中心,找各属性值有明显减少趋势的点。

根据热力学定律,在温度为 t 的情况下,能量差所表现的概率如下: $P(\Delta E) = \exp(-\Delta E/kt)$, k 是麦克斯韦常量。转换到模拟退火算法,则变成: $P = \exp(-c/t) > r$, c 是评估函数的差, r 是 0-1 之间的乱数。假设所求解的问题是目标函数最小化问题 $\Delta f = f(x') - f(x)$,若 $\Delta f < 0$,则透过概率函数接受 $f(x')$ 为新解。接着判断是否满足降温条件,若满足,则通过冷却机制降温,反之,维持目前温度 $T = \alpha \times T, \alpha \in [0, 1]$ 。之后再判断是否达到终止条件,例如当达到设定的叠代次数或者是连续几次叠代目前解都不再改变时,则停止退火。

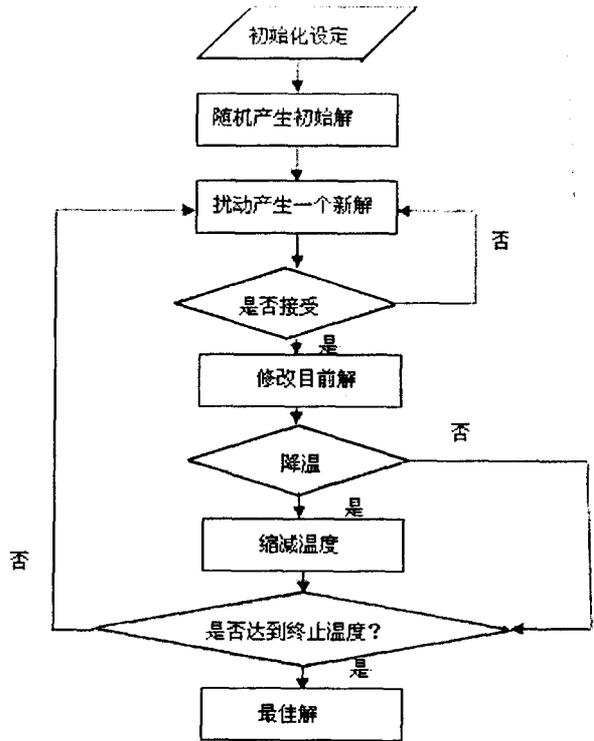


图 4 模拟退火流程图

模拟退火过程需要注意的一些问题:

初始温度要够高才能移动到任何的状态,但是又不能太高,否则会导致在一段时间内皆用乱数在凑解答。在油藏边缘的模拟中寻找的初始点应避开含油丰富的油藏中心位置。

模拟退火的最终温度通常是零,但这样会耗掉许

多模拟时间。温度接近于零,其周遭状态近乎是一样的。所以寻找一个低到可接受的温度,既可以节省模拟时间,又不影响最终结果。在文中,尽可能寻找一个可接受的外部轮廓结构图,不一定要和实际轮廓完全吻合。模拟出来的油藏边缘轮廓只不过是视觉需求,最终显示的还是真实的轮廓图。所以,只要是油藏的含油量减少到一定的程度,就可以退火结束。

每次降低温度的差距以及在同一温度反复搜寻最适合的解会导致指数般增长的搜寻空间。低温需要较多反复次数以避免找到局部最大值,但高温需要的反复次数相对较少。

应用在油藏边缘模拟中,每次在 x, y, z 轴上向外扩散的幅度不易过大或过小,过大的话虽然反复次数少,但是模拟出来的轮廓误差较大;过小的话虽然最终解准确,但是多次叠代需要较长的时间。

模拟退火法以扰动的机制来产生一个扰动解,再以概率函数判断是否接受此扰动解为叠代的新解;若不被接受,就再以扰动重新产生一个扰动解,并以概率函数重新判断。每代重覆以上的步骤,直到接受此次叠代的新解为止。扰动过程如图 5 所示。

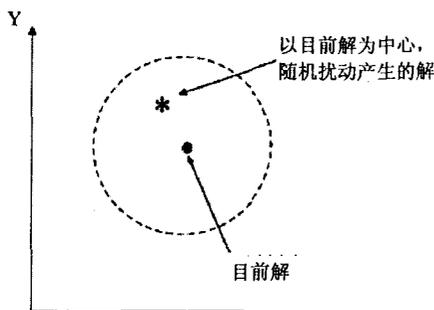


图 5 扰动过程举例

模拟退火法利用概率函数有机率的接受较差的扰动解为新解^[8],使其避免了传统梯度搜索法往往陷入区域解的缺点,而使模拟退火法有机会跳脱区域解,往全域最佳解收敛。一般的概率函数方程式如下:

$$p(X') = \begin{cases} 1, \Delta f \leq 0 \\ \exp(-\Delta f/T), \Delta f > 0 \end{cases}$$

$$\Delta f = f(X') - f(X)$$

T 为目前温度。当 $\Delta f \leq 0, p(X') = 1$, 在 $\Delta f > 0$ 时会随机产生一个介于 0 到 1 的一个小于 1 大于 0 的值。随机值 r 与 $p(X')$ 比较,若 $p(X') \geq r$ 则接受扰动解 $f(X')$;反之,则不接受。当 T 越高或越小时,则 $p(X')$ 越大,相对的扰动解被接受成新解的概率越高。因此 T 会随着叠代次数的增加而逐渐下降,所以较差的扰动解被接受成新解的机会也会随着 T 的下降而越来越

小。所以当叠代到最后因为温度 T 已到达低点,这时系统只会接受较佳的扰动解为新解。而扰动解 $f(X')$ 若小于目前解 $f(X)$ 则一定接受为新解,但若是 $f(X') > f(X)$ 则接受为新解的概率随着 Δf 的变大而越小。

文中,参数是指油藏的含油、沉积、成岩、流体等 44 个属性。为了保证最终解的准确性,目前解的选取应该尽量避开油藏中心,找含油量有明显减少趋势的点。同时,需要找 44 个目前解,扰动产生 44 个最终解。由于最终解是由不同属性的变化产生的,难免存在片面性,所以还要通过遗传算法,以模拟退火算法产生的 44 个最终解为个体,找到最准确的油藏模型。

遗传算法是模拟达尔文的遗传选择和自然淘汰的生物进化过程的计算模型,是一种通过模拟自然进化过程搜索最优解的方法,是一类可用于复杂系统优化的具有鲁棒性的搜索算法,借鉴生物学中的染色体和基因的概念,模仿自然界生物的遗传和进化机理,直接以适应度作为搜索信息,无需导数等其它辅助信息。遗传算法使用概率搜索技术,而非确定性规则。

文中应用的遗传算法的搜索空间 U 为一固定值,即为遗传算法产生的 40 个最终解,也就是遗传算法中的初始种群。

4.2 双向链表存储

与其他数据存储形式相比,链表在程序运行的过程中动态地分配或者回收存储空间的分配内存,可以很大程度上提高空间利用率。双向链表的一个显著的优点就是避免了在插入/删除操作时对单元搜索,有很高的效率。双向链表不但可以由首(尾)地址单元向前、向后搜索,而且还可以从中间的某一单元开始,分别向两边搜索。在油田项目中,采用双向链表存储的方式,将数据表中的油藏边缘数据存储到双向链表中(见图 6),同时,采用数组来存储每条双向连表的表头指针,虽然数组的这种静态存储不能及时地放空间,但是顺序表可以便于快速搜索到数据单元,均衡时间效率和空间效率,双向链表的头地址在这里使用顺序表存储最为合适。

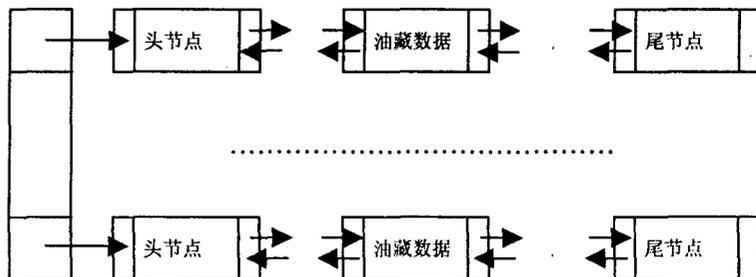


图 6 边缘数据的双向链表存储结构

秒一次的采样来说是足够的,完全满足实时性的要求。在中断子程序中实时调用 PID 控制算法的程序如下:

```

ADCINT ISR;高优先级模式的 ADC 中断子程序
CLRC SXM ;抑制符号位扩展
LDP #4 ;指向用户变量区
LAR AR2, # RESULT0;把当前 A/D 转换的结果存储到 AR2 中
MAR *, AR2 ;把 AR2 设置为当前辅助寄存器
LACC *, 12 ;由于 DSP 内嵌的 A/D 转换器是 12 位的,所以左移 10 位加载到 ACC 高位
SACH ADRESULT;变量 ADRESULT 存储当前 A/D 转换的结果
CALL PID Control;调用数字 PID 控制子程序,保证实时控制
CLRC INTM ;开总中断,因为一进中断就自动关闭总中断
RET ;中断返回

```

3 试验结果

数字电子负载的功能除了完成作为负载的功能以外,还要实现测量的任务,从测量功能上看,要测量 3 个参数:输出的直流电压、直流电流和输出的纹波电压,这是 3 个性质完全不同的电信号。电子负载工作模式选用恒压模式,连接市售输出为 5V/0.48A 直流电源,设定为 0.51A,0.32A,0.15A 吸取导入电流,测试电子负载的电流实时调节性能,由图 4 测试结果可知:调整负载电流时输出电压很稳定,负载电流精度控制在 4mA 左右,电流调节过程中电源输出电压的变化很小,保持在 5V 左右,切换电流档位时出现的电压峰值为 5.26V,仍在允许的范围内。采用数字电子负载对负载电流进行数字化的实时控制调节,可以根据电源供应器的特性,自动地调节和控制负载电流大小及其上升/下降率,以适应种类繁多的电源供应器,不会

引起电源供应器的输出电压和输出电流的震荡冲击^[7,8]。

4 结束语

目前国内广泛使用的直流电子负载负载仪对负载电流的调节和控制采用硬件模拟电路的方法,调节和控制的适应性和实时性差,仅仅具有静态负载的测试功能,不能适应不同的电源供应器的具体情况。经试验测试,利用 TMS320F2812 作为控制器的电子负载,从硬件上保证了 PID 控制策略的实时性,特别是适应了负载的动态瞬变,实现了在两个不同负载准位之间的动态切换,具有了动态负载的测试功能,可以满足一些较高要求和特殊测试的场合。

参考文献:

- [1] Zhang Wanfeng. DSP implementation of predictive control strategy for power factor correction[J]. APEC, 2004(2): 67 - 73.
- [2] 刘 琛,梁昔明.一种结合数值优化的 PID 制器韵设计与仿真[J]. 计算机技术与发展,2007,17(4):24 - 26.
- [3] 史增树,张建红. TMS320F2812 在全数字开关电源中的应用[J]. 现代电子技术,2006(13):97 - 99.
- [4] Texas Instruments. TMS320X281X Serial Peripheral Interface (SPI) Reference Guide[M]. Texas, USA: TI, 2006.
- [5] 张 琴. 嵌入式目标模块在 DSP 系统开发中的应用[J]. 华中科技大学学报,2005(2):96 - 98.
- [6] 严 杰,王 莉,王志强. 基于 Matlab 的功率 MOSFET 建模[J]. 电力电子技术,2005(3):23 - 26.
- [7] User's Guide Agilent Technologies DC Electronic Loads Models N33045A[M]. USA: Agilent, 2005.
- [8] 洪贞启. N3305A 电子负载与电源检定[J]. 电子测试,2008(4):74 - 76.

(上接第 240 页)

5 结束语

整个百万节点油藏模型系统建立在对数据详细分析处理的基础上,硬件资源消耗少,视觉方面显示快。不但有效地提高了虚拟场景实时渲染速度,而且通过边缘渲染的方法提高了视觉效果。

参考文献:

- [1] 岳大力. 基于三维储层构型模型的油藏数值模拟及剩余油分布模式[J]. 中国石油大学学报,2008,4(2):21 - 27.
- [2] Gao Pengdong. Large data processing based on simplification and subdivision[C]//Proceedings of SPIE - The International Society for Optical Engineering. Fourth International Conference on Virtual Reality and Its Applications in Industry. [s.

- [1]:[s. n.],2004:35 - 41.
- [3] 姜英姿. 大规模数据的计算机处理技术[J]. 徐州工程学院学报,2005,10(5):13 - 15.
- [4] 王书海. 基于 SQL Server 应用系统的大数据量实时处理技术[J]. 实验室研究与探索,2008(3):16 - 18.
- [5] 吴 磊. GIS 大数据量编辑处理中快速捕捉的优化策略[J]. 武汉理工大学学报,2005,4(2):315 - 318.
- [6] 程 燕. 图像超分辨率重建关键技术的研究[D]. 上海:上海交通大学,2007:39 - 40.
- [7] 陈学工. 模拟退火遗传算法的轮廓线拼接研究[J]. 计算机仿真,2009(2):208 - 211.
- [8] Sheldon M. Ross 统计模拟[M]. 第 4 版. 北京:人民邮电出版社,2007.