(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 105940402 B
(45)授权公告日 2019.01.29

(21)申请号 201580004212.X

(22)申请日 2015.01.09

(65)同一申请的已公布的文献号
申请公布号 CN 105940402 A

(43)申请公布日 2016.09.14

(30)优先权数据
1450176 2014.01.10 FR

(85)PCT国际申请进入国家阶段日
2016.07.11

(86)PCT国际申请的申请数据
PCT/FR2015/050051 2015.01.09

(87)PCT国际申请的公布数据
WO2015/104512 FR 2015.07.16

(73)专利权人 IMA生物科技公司
地址 法国罗斯

(72)发明人 法比安·帕默拉尔
乔纳森·M·斯陶贝尔

(74)专利代理机构 中原信达知识产权代理有限责任公司 11219
代理人 鲁山 孙志湧

(51)Int.Cl.
G16B 45/00(2019.01)
G16B 50/00(2019.01)

(56)对比文件
CN 101782930 A,2010.07.21,
US 2012258485 A1,2012.10.11,
WO 2009156897 A1,2009.12.30,
CN 1668921 A,2005.09.14,

审查员 吴瑶

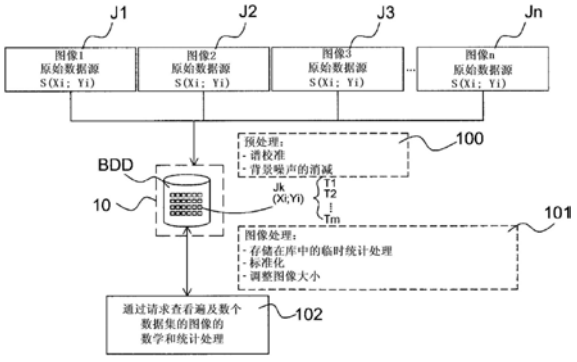
权利要求书3页 说明书7页 附图3页

(54)发明名称

用于处理分子成像数据的方法及对应的数据服务器

(57)摘要

本发明主要涉及用于处理多个谱数据集(J1-Jn)的方法或用于记录多个谱数据集(J1-Jn)的方法,所述谱数据集(J1-Jn)旨在供分子成像方法使用,每个谱数据集(J1-Jn)由一组空间位置(Xi,Yj)定义,所述空间位置中的每一个与具有包含一组分子信息的至少两个维度的分子谱(S(Xi,Yj))相关联,其特征在于所述方法尤其包括以下步骤:对于每个数据集(J1-Jn),将与每个位置(Xi,Yj)相关联的所述分子谱切成多个谱分段(T1-Tm);将针对每个数据集(J1-Jn)的每个位置(Xi,Yj)所获得的所述分段(T1-Tm)插入到数据库(BDD)中;按照对于感兴趣的分子信息的请求在所述数据库(BDD)中选择包含所述感兴趣的分子信息的一个或多个分段(T1-Tm);以及在每个分段(T1-Tm)内选择所述感兴趣的分子信息。



CN 105940402 B

1.一种用于处理多个谱数据集J1-Jn的方法,所述谱数据集J1-Jn旨在供分子成像过程利用,每个谱数据集J1-Jn由一组空间位置(Xi,Yj)定义,其中所述空间位置中的每一个与包含一组分子信息的至少两个维度的分子谱S(Xi,Yj)相关联,其特征在于所述方法尤其包括以下步骤:

-对于每个数据集J1-Jn,将与每个位置(Xi,Yj)相关联的所述分子谱切成谱的数个部分T1-Tm,每个所述部分包含所述谱的所述分子信息中的一些,

-将针对每个数据集J1-Jn的每个位置(Xi,Yj)所获得的所述部分T1-Tm插入到数据库BDD中,使得谱的索引部分T1-Tm的集合与每个数据集J1-Jn的每个位置(Xi,Yj)相关联,

-按照对于感兴趣的分子信息的请求在所述数据库BDD中选择包含所述感兴趣的分子信息的所述部分T1-Tm,以及

-在每个部分T1-Tm内选择所述感兴趣的分子信息,

其中,所述方法进一步包括以下步骤:

-以至少一个预定间距P'将所述分子谱S(Xi,Yj)切成谱的部分T1-Tm,

-插入参考轴Aref,所述参考轴Aref建立了不同部分T1-Tm的点索引与所对应的分子信息之间的对应性,以及

-根据所述参考轴Aref和所述预定间距P'来选择包含所述感兴趣的分子信息的所述部分T1-Tm。

2.根据权利要求1所述的方法,其特征在于所述方法包括以下步骤:以一组数据图C1-Cn的形式查看所述感兴趣的分子信息,所述一组数据图中的每一个对应于数据集J1-Jn。

3.根据权利要求1所述的方法,其特征在于,在将部分T1-Tm插入到所述数据库BDD之前或之后,所述方法包括由谱校准和/或背景噪声消减和/或数据集J1-Jn内标准化组成的至少一个预处理步骤。

4.根据权利要求1所述的方法,其特征在于所述方法包括以下步骤:调整具有不同空间大小的所述数据集J1-Jn的所述空间位置(Xi,Yj)的大小,以便使得所述空间位置(Xi,Yj)与具有最细空间大小的所述数据集J1-Jn相校准。

5.根据权利要求1所述的方法,其特征在于所述方法还包括提取所有所述分子谱的感兴趣的每个峰的数据集的步骤。

6.根据权利要求5所述的方法,其特征在于,所提取的数据集从由以下各项构成的组中选择:最大强度Pk1-Pk1、平均强度、峰下面积、和信噪比。

7.根据权利要求5所述的方法,其特征在于,根据由信噪比和/或谱分辨率的质量准则所定义的峰选择准则来执行提取。

8.根据权利要求1所述的方法,其特征在于所述方法包括以下步骤:使多个所述数据集J1-Jn的所述分子谱相互标准化,以便将不同的所述数据集J1-Jn进行相互比较。

9.根据权利要求8所述的方法,其特征在于为了执行所述标准化步骤,所述方法包括以下步骤:

-选择为所有所述数据集J1-Jn所共有的一个分子或数个内生或外生参考分子Mref,

-对于每个数据集J1-Jn,根据这个或这些参考分子Mref来计算标准化因子Fnor,以及

-使用所述标准化因子校正所述感兴趣的分子信息以获得一组标准化的数据集J1-Jn。

10.根据权利要求9所述的方法,其特征在于,根据存在于与所述数据集J1-Jn对应的一

组样本上或者在与所述数据集J1-Jn对应的所述样本的特定参考区域Z中的至少一个参考分子Mref来定义所述标准化因子Fnor。

11.根据权利要求8所述的方法，其特征在于所述方法包括以下步骤：以多个数据图C1'-Cn'的形式查看、比较并分析具有为所有所述图所共有的色标的所述感兴趣的分子信息，所述多个数据图中的每一个来自标准化的数据集J1-Jn。

12.根据权利要求11所述的方法，其特征在于，所共有的色标基于感兴趣的分子的所有所述数据集J1-Jn的最小强度和最大强度。

13.根据权利要求1所述的方法，其特征在于，所述数据集J1-Jn是通过质谱分析过程获得的。

14.根据权利要求1所述的方法，其特征在于，所述数据集J1-Jn是通过正电子发射断层扫描PET或磁共振成像MRI类型的成像过程获得的。

15.一种数据服务器(10)，所述数据服务器(10)包括存储软件指令的存储器，所述软件指令用于执行根据权利要求1中所定义的所述用于处理多个数据集J1-Jn的方法的步骤。

16.一种用于在分子成像过程期间记录多个谱数据集的方法，每个数据集由一组空间位置定义，所述空间位置中的每一个与包含一组分子信息的至少两个维度的分子谱相关联，其特征在于所述方法包括以下步骤：

-在谱数据的采集期间，并且对于每个谱数据集，将与每个位置相关联的所述分子谱切成谱的数个部分，每个部分包含所述谱的所述分子信息中的一些，

-将针对每个数据集的每个位置所获得的所述部分插入到数据库中，使得谱的索引部分的集合与每个数据集的每个位置相关联，并且可选地

-按照对于感兴趣的分子信息的请求在所述数据库中选择包含所述感兴趣的分子信息的所述部分，以及

-在每个部分内选择所述感兴趣的分子信息，

其中，所述方法进一步包括以下步骤：

-以至少一个预定间距P'将所述分子谱S(Xi,Yj)切成谱的部分T1-Tm，

-插入参考轴Aref，所述参考轴Aref建立了不同部分T1-Tm的点索引与所对应的分子信息之间的对应性，以及

-根据所述参考轴Aref和所述预定间距P'来选择包含所述感兴趣的分子信息的所述部分T1-Tm。

17.根据权利要求16所述的方法，其特征在于所述方法包括以下步骤：以一组数据图C1-Cn的形式查看所述感兴趣的分子信息，所述一组数据图中的每一个对应于数据集J1-Jn。

18.根据权利要求16所述的方法，其特征在于，在将部分T1-Tm插入到所述数据库BDD之前或之后，所述方法包括由谱校准和/或背景噪声消减和/或数据集J1-Jn内标准化组成的至少一个预处理步骤。

19.根据权利要求16所述的方法，其特征在于所述方法包括以下步骤：调整具有不同空间大小的所述数据集J1-Jn的所述空间位置(Xi,Yj)的大小，以便使得所述空间位置(Xi,Yj)与具有最细空间大小的所述数据集J1-Jn相校准。

20.根据权利要求16所述的方法，其特征在于所述方法还包括提取所有所述分子谱的

感兴趣的每个峰的数据集的步骤。

21.根据权利要求20所述的方法，其特征在于，所提取的数据集从由以下各项构成的组中选择：最大强度Pk1-Pk1、平均强度、峰下面积、和信噪比。

22.根据权利要求20所述的方法，其特征在于，根据由信噪比和/或谱分辨率的质量准则所定义的峰选择准则来执行提取。

23.根据权利要求16所述的方法，其特征在于所述方法包括以下步骤：使多个所述数据集J1-Jn的所述分子谱相互标准化，以便将不同的所述数据集J1-Jn进行相互比较。

24.根据权利要求23所述的方法，其特征在于为了执行所述标准化步骤，所述方法包括以下步骤：

-选择为所有所述数据集J1-Jn所共有的一个分子或数个内生或外生参考分子Mref，

-对于每个数据集J1-Jn，根据这个或这些参考分子Mref来计算标准化因子Fnor，以及

-使用所述标准化因子校正所述感兴趣的分子信息以获得一组标准化的数据集J1-Jn。

25.根据权利要求24所述的方法，其特征在于，根据存在于与所述数据集J1-Jn对应的一组样本上或者在与所述数据集J1-Jn对应的所述样本的特定参考区域Z中的至少一个参考分子Mref来定义所述标准化因子Fnor。

26.根据权利要求23所述的方法，其特征在于所述方法包括以下步骤：以多个数据图C1'-Cn'的形式查看、比较并分析具有为所有所述图所共有的色标的所述感兴趣的分子信息，所述多个数据图中的每一个来自标准化的数据集J1-Jn。

27.根据权利要求26所述的方法，其特征在于，所共有的色标基于感兴趣的分子的所有所述数据集J1-Jn的最小强度和最大强度。

28.根据权利要求16所述的方法，其特征在于，所述数据集J1-Jn是通过质谱分析过程获得的。

29.根据权利要求16所述的方法，其特征在于，所述数据集J1-Jn是通过正电子发射断层扫描PET或磁共振成像MRI类型的成像过程获得的。

30.一种数据服务器(10)，所述数据服务器(10)包括存储软件指令的存储器，所述软件指令用于执行根据权利要求16中所定义的所述用于处理多个数据集J1-Jn的方法的步骤。

用于处理分子成像数据的方法及对应的数据服务器

技术领域

[0001] 本发明涉及基于使用大容量数据库结构以及基于对应数据的服务器来处理分子成像数据的过程。分子成像通常涵盖通过尽可能少的侵入性手段或尽可能少地破坏被观察的生物或生物物理系统从根本上观察活体外或活体内的器官和/或有机体的分子手术的所有成像技术。

背景技术

[0002] 分子成像正在许多技术的演进中尤其在质谱成像(MSI)、磁共振成像(MRI)或拉曼光谱法中,经历相当大的上升。这些技术被用来指导对诸如杀虫剂、药物、蛋白质或脂类的定向内生或外生化合物的生物分布的研究,以便研究它们在生物系统中的作用。随着来自软件的竞争,这些技术的应用还搜索潜在的非监视生物标记。

[0003] 但是,在分子成像中待分析和解释的数据的增加对数个数据集的综合分析构成了限制,该数个数据集中的每个与图像相关联。数据正变得难以用于综合表示或者在统计上进行分析。

[0004] 确实,数据集信息的大容量常常大于常规计算机的随机存取存储器的大小(与超级计算机对照)。例如利用质谱成像,50,000个位置的图像不再能够将所有原始信息存储在存储器中并且因此需要执行计算以减少对存储器的影响并且允许利用这些图像。计算减少数据并且因此由于不考虑所有可用的信息而导致了偏差。并且,当前的分析工具没有使数个数据集相对于彼此标准化的手段。

[0005] 存在主要与数据集的采集机器人的制造商相关联的许多存储格式。然而,这些格式当中均未被适配和优化以供细致且迅速地查询。分子成像的一些领域已强调了存储格式,诸如在MRI(磁共振成像)或质谱成像中使用的Analyze 7.5。质谱成像还包含却不构成用于处置成像数据的存储系统的格式iMZML。

[0006] 基于数据的分层结构的HDF5(“分层数据格式5”)类型的格式近来已被用于在质谱成像中存储大容量数据。这些格式经由互联网导航构建了远程查询接口。它们与利用非分层文件的格式相比使得能够实现更快的统计计算。然而这些格式在研究数个数据集时不同时产生数据的全局比较和可搜索分析。

[0007] 因此需要如下过程:在信息不丢失的情况下,相对于彼此分析、比较并查询数个数据集以便避免引入分析偏差。

发明内容

[0008] 本发明的目的在于通过提出用于处理多个谱数据集的过程来有效地应对这种需要,所述谱数据集尤其旨在供分子成像过程利用,每个数据集由一组空间位置定义,其中所述空间位置中的每一个与包含一组分子信息的至少两个维度的分子谱相关联,其特征在于所述过程尤其包括以下步骤:

[0009] -对于每个数据集,将与每个位置相关联的分子谱切成分子信息的数个部分或谱

的数个部分，每个部分包含分子信息的缩减集，

[0010] -将针对每个数据集的每个位置所获得的部分插入到数据库中，使得谱的索引部分的集合与每个数据集的每个位置相关联，

[0011] -按照对于感兴趣的分子信息的请求在数据库中选择包含感兴趣的分子信息的部分，以及

[0012] -在每个部分内选择所述感兴趣的分子信息。

[0013] 本发明还涉及用于在分子成像过程期间记录多个谱数据集的过程，每个数据集由一组空间位置定义，具有包含一组分子信息的至少两个维度的分子谱与所述空间位置中的每一个相关联，其特征在于所述过程包括以下步骤：

[0014] -在谱数据的采集期间，并且对于每个谱数据集，将与每个位置相关联的分子谱切成谱的数个部分，每个部分包含所述谱的分子信息中的一些，

[0015] -将针对每个数据集的每个位置所获得的部分插入到数据库中，使得谱的索引部分的集合与每个数据集的每个位置相关联，并且可选地

[0016] -按照对于感兴趣的分子信息的请求在数据库中选择包含感兴趣的分子信息的部分，以及

[0017] -在每个部分内选择所述感兴趣的分子信息。

[0018] 更普遍地，本发明提出了对数据集进行处理，所述数据集以通过将与构成所述数据集的空间位置的集合中的每一个空间位置相关联的分子谱切成部分的自动的方式被有利地采集。与每个谱对应的是谱的连续部分，所有部分包括所关联的谱的全部分子信息。根据本发明，能够在谱数据的采集期间或者在谱数据的分析期间以自动的方式有利地进行切成部分的步骤。

[0019] 在第一实施例中连续部分不重叠并且仅相邻，使得来自相关谱的分子信息不丢失或不冗余。在另一实施例中相关谱的一些部分可以部分地重叠，使得一些分子信息可以存在于同一谱的两个连续谱部分中。

[0020] 本发明根本上将待分析的数据的加载周期减少到了能够通过选择部分来恢复数据库中的有用信息的程度。还能够在数据的采集期间当数据被记录时即刻切割谱数据，使得谱数据的部分直接在样本的分析期间通过分子成像过程被集成。

[0021] 由于将信号切成了部分，在不减少谱的情况下，本发明就在为观察其整体的生理事件的相同分析中的内生或外生分子的药物动力、药效以及定量的组合研究方面而言，在不减少信息的情况下也工作，所述组合研究诸如为筛选方法(数种药物候选的分配比较)、定量的方法、受体或组织筛查的占用研究的方法、新陈代谢或3D重建研究。本发明还优化了对数个数据集的分子数据成像的统计处理和表示所必需的存储器的使用。

[0022] 根据实施例，本发明的用于数据处理或记录的过程包括以下步骤：以一组数据图的形式查看感兴趣的分子信息，数据图中的每一个对应于数据集。

[0023] 有利地，根据确定间距来切去分子谱，所述间距可以在谱间相同或不同。

[0024] 根据实施例，本发明的用于数据处理或记录的过程包括以下步骤：

[0025] -根据至少一个预定间距将所有分子信息切成谱的部分，

[0026] -插入参考轴，所述参考轴建立了谱的不同部分的点索引与对应的分子信息之间的对应性，以及

[0027] -根据参考轴和预定间距来选择包含感兴趣的分子信息的谱部分。

[0028] 根据实施例，在将部分插入到数据库中之前或之后，所述过程包括由谱校准和/或背景噪声消减和/或数据集内标准化组成的至少一个预处理步骤。

[0029] 根据实施例，本发明的用于数据处理或记录的过程包括以下步骤：调整具有不同空间大小的数据集的空间位置的大小，以便使得空间位置与具有最细空间大小的数据集相校准。

[0030] 根据实施例，本发明的用于数据处理或记录的过程还包括提取数据集的步骤，诸如提取所有分子信息的感兴趣的每个峰的最大强度、平均强度、峰下面积、信噪比。

[0031] 根据实施例，根据由信噪比和/或谱分辨率的质量准则所定义的峰选择准则来实现提取。

[0032] 根据实施例，本发明的用于数据处理或记录的过程包括以下步骤：使多个数据集的分子信息的集合相互标准化，以便然后将不同的数据集进行相互比较。

[0033] 根据实施例，为了着手标准化步骤所述过程能够包括以下步骤：

[0034] -选择为所有数据集所共有的一个参考分子或数个内生或外生参考分子，

[0035] -对于每个数据集，根据这个或这些参考分子来计算标准化因子，以及

[0036] -使用这个标准化因子校正感兴趣的分子信息以获得一组标准化的数据集。

[0037] 根据实施例，根据存在于与数据集对应的一组样本上或者在与数据集对应的样本的特定参考区域中的至少一个参考分子来定义标准化因子。

[0038] 根据实施例，本发明的用于数据处理或记录的过程包括以下步骤：以多个数据图的形式查看、比较并分析具有为所有图所共有的色标的感兴趣的分子信息，多个数据图中的每一个来自标准化数据集。

[0039] 根据实施例，所共有的色标基于感兴趣的分子的所有数据集的最小强度和最大强度。

[0040] 根据实施例，所述数据集是通过质谱分析过程获得的。

[0041] 根据实施例，所述数据集是通过正电子发射断层扫描(PET)或磁共振成像(MRI)类型的成像过程获得的。

[0042] 有利地，根据本发明的记录和/或处理过程步骤中的全部或一些是自动的。

[0043] 本发明的另一目的是数据服务器，所述数据服务器包括存储软件指令的存储器，所述软件指令用于执行根据本发明的所述用于处理多个数据集的过程的至少一些步骤。

[0044] 本发明的另一目的是可由计算机读取的数据介质，所述数据介质包括指令，所述指令可由计算机执行并且被适配成允许计算机系统执行根据本发明的所述用于处理或者记录谱数据的过程的至少一个步骤。

[0045] 因此，本发明提出了计算机程序，所述计算机程序包括用于在所述程序在计算机上运行时执行上面所说明的步骤中的全部或一些的程序代码指令。

[0046] 有利地，计算机程序包括用于执行至少将分子谱切成谱的部分的步骤的程序代码指令。

附图说明

[0047] 从以下描述以及对附图的检验将更清楚地理解本发明，仅通过图示但非不限制本

发明的方式给出附图。

[0048] 图1示意性地图示了用于获得由根据本发明的过程所利用的数据集的质谱分析过程的不同步骤；

[0049] 图2是根据本发明的用于执行处理数据的过程的集成数据库管理系统的概况图；

[0050] 图3是用于获得与数据集的每个位置相关联的谱部分的步骤的示意图示；

[0051] 图4示出了用于建立分子信息与所对应的谱部分之间的对应性的步骤；

[0052] 图5示意性地图示了用于精确地比较数个图像的图像间标准化步骤。

[0053] 相同、类似或模拟元件在附图间保持相同的参考标记。

具体实施方式

[0054] 针对MALDI类型的质谱分析方法来描述用于执行根据本发明的过程的以下示例。当然，能够本质上相同地使用除MALDI以外的成像装置，诸如例如这些源：SIMS、DESI、LAESI、DIOS、ICP、显微镜MALDI、SNOM、SMALDI、LA-ICP、ESI(组织上的液体提取)、MILDI、JEDI、ELDI等。该过程还适用于产生可由诸如例如PET(“正电子发射断层扫描”)或MRI(磁共振成像)的分子成像的方法利用的数据的任何其它方法。

[0055] 图1示意性地图示了由MALDI质谱分析过程执行的不同步骤。第一步骤E1包括通常通过冷冻切片来切去组织1的薄片，组织1的薄片然后被放置在载片2上。在步骤E2中离子化基质3的细均匀层然后被沉积在组织1的薄片上。采集步骤E3包括生成具有预定大小(例如每采集区域大约100微米)的自动激光脉冲4，以使基质3的分子离子化。以这种方式离子化的分子由质谱分析仪如其本身已知的那样进行分析。能够参考文档FR2973112以得到关于执行该过程的更多细节。

[0056] 产生了由一组空间位置(Xi,Yj)所定义的数据集Jk，与该组空间位置中的每一个相关联的是一组分子信息。空间位置被定义为参考X轴和Y轴的标记的函数。所有分子信息在这种情况下尤其由谱S形成，谱S具有将强度指示为分子质量的函数的两个维度。例如，所获得的数据集Jk遍及20,000个空间位置(Xi,Yj)具有每谱50,000个强度(或50,000个点)。根据本发明的过程基于处理多个数据集J1-Jn。

[0057] 如图3中所示，对于很可能将以图像的形式示出的每个数据集Jk，与每个位置(Xi,Yj)相关联的分子谱S(Xi,Yj)被切成谱的数个部分T1-Tm。部分T1-Tm是按照预定间距P’来切去的。例如，对于每空间位置具有50,000个点的每个谱，能够按照10,000个点的间距P’切割每个谱。或者，谱S(Xi,Yj)被切成五个部分T1-T5，每个部分具有10,000个点。作为变例，所使用的间距P’能够是可变的。

[0058] 还定义了图4中示出的参考轴Aref，从而建立了不同部分T1-Tm的位置索引与所对应的分子质量之间的对应性。

[0059] 在插入到数据库BDD中之前，该过程优选地经由模块100执行谱S(Xi,Yj)的至少一个预处理步骤。这个预处理步骤能够包括谱校准。出于这个目的，考虑参考峰并且偏移谱，以使得在这些参考峰上定位对应峰的值。还可以通过根据谱的最小值与谱的最小值将必须与其对应的参考值之间的距离重新调整谱，来执行背景噪声的消减。

[0060] 此外，能够(通过内部标准、通过使用强度的总数...)执行不同类型的信号标准化(内部标准化)。换句话说，在这里执行数据集内的标准化，即独立于其它数据集而对每个数

据集进行标准化，其对应于为技术人员本身所已知的步骤。这个数据集内的标准化作为在下文中详述的数据集J1-Jn之间的后续标准化步骤的补充(数据集间标准化)被执行，数据集J1-Jn之间的后续标准化步骤将精确且安全地将数据集J1-Jn进行相互比较。

[0061] 并且，该过程包括提取数据集的步骤，诸如提取谱S(Xi,Yj)的感兴趣的每个峰的最大强度Pk1-Pk1、平均强度、峰下面积、信噪比。能够根据被称作“峰拾取”的方法来提取谱S(Xi,Yj)的每个峰Pk1-Pk1的最大强度。根据由信噪比和/或谱分辨率的质量准则所定义的峰选择准则来实现提取。

[0062] 对每个数据集J1-Jn的每个位置(Xi,Yj)以及参考轴Aref和先前所提取的峰Pk1-PK1所获得的部分T1-Tm被插入到如图2中所示出的数据库BDD中。以这种方式，索引部分T1-TM的集合和峰Pk1-PK1的集合以及上述补充数据(每个峰的平均强度、面积)与每个数据集J1-Jn的每个位置(Xi,Yj)相关联。

[0063] 数据库BDD是使得能够实现插入数据的查看、管理、解释和统计分析的NoSQL(“Not only Structured Query Language”)或SQL类型的大容量数据库。例如能够从以下非穷尽的系统中选择数据库管理系统：Hypertable、Haddoop、Cassandra、MongoDB、PolyBase、Hadoop on Azure、Hive、Pig。数据库系统确保数据的相干性、可靠性和切合性。用于复制和保存的技术能够确保信息的可靠性。

[0064] 下面描述用于在数据库BDD中选择切合信息的步骤。按照对于感兴趣的分子信息的请求，选择包含所请求的分子信息的数据集J1-Jn的集合的位置的部分。出于这个目的，对于每个数据集J1-Jn的每个位置(Xi,Yj)，根据参考轴Aref以及所请求的分子信息来选择包含所请求的分子信息的部分Tp。

[0065] 因此，图4图示了根据五个点的间距P’将谱切成部分，以便更容易理解这个步骤。对于起源于质谱分析方法的谱来说情况常常是这样的，分子质量的标度不是线性的。在用户需要关于例如具有值3m/z的分子质量的分子的信息的情况下，参考轴Aref指明与分子质量的这个值对应的点的索引是9，使得能够借助于间距P’推断出包含感兴趣的分子质量的强度的部分Tp是第二部分T2。

[0066] 替换地，能够使用存储分子质量的最小端点和最大端点以及每个对应部分T1-Tm的点的指数化的表。在这种情况下，对于请求的分子信息，根据这些最小端点和最大端点来选择部分T1-Tm的标识符。

[0067] 在为所有数据集J1-Jn中的每一个点选择的每个部分T1-Tm内，然后选择所请求的分子质量的对应强度。然后成为可能的是，以一组数据图C1-CN为形式查看感兴趣的分子存在于不同图C1-CN上的强度的分布，数据图C1-CN中的每一个对应于数据集J1-Jn。颜色被认为是分子信息关于色标的强度的函数。替换地，或者通过补充，峰Pk1-PK1的提取迅速地产生了提供对感兴趣的分子的强度分布的初始分析的图像。

[0068] 该过程优选地包括为随后将数个数据集进行相互比较而使数据集J1-Jn的分子信息标准化的步骤(参考模块101)。出于这个目的，选择为所有数据集J1-Jn所共有的参考分子Mref。替换地，选择为所有数据集J1-Jn所共有的数个内生或外生分子。能够认为参考分子Mref存在于完整样本中或者更具体地在对应样本的参考区域中。

[0069] 更准确地说，对于每个数据集J1-Jn，标准化因子Fnor是根据这个或这些参考分子Mref来计算的。然后使用这个标准化因子Fnor校正感兴趣的分子信息，以产生标准化集J1-

Jn的数据集。

[0070] 例如,对于与图5中所示出的图C1-C3对应的三个数据集,参考分子Mref＝500m/z被选择为标准化分子(标准化标准)。这个分子Mref在图C1-C3的区域Z中分别具有5、10和15的值。对于每个图C1-C3,基于Mref的这些值中的每一个来计算标准化因子Fnor:

[0071] Fnor＝C1-C3的Mref值的平均值/相关图的Mref值。

[0072] 在图5中所示出的示例中,对于图C1、C2和C3分别获得了2、1和10/15的标准化因子Fnor。然后这个系数被应用于与所有数据集J1-Jn的不同位置相关联的谱的所有强度值,以及特别是在位置Xi;Yj处的感兴趣的分子Mi＝300m/z的强度值。获得了能够直接相互比较的标准化图C1'-C3'。

[0073] 标准化校正了与在实验条件下由于样本的制备及其分析而导致的改变相关联的强度。实际上,在前面的示例中,显然将一直能够达成感兴趣的分子Mi＝300m/z较强存在于与图C3对应的样本中的错误结论,尽管在应用标准化因子Fnor之后,研究揭示了感兴趣的分子Mi均匀存在于在位置(Xi,Yj)处具有强度为200的三个样本中。

[0074] 按照这个标准化,能够以标准化的数据图C1'-C3'的形式查看、比较并分析具有为所有图C1'-C3'所共有的色标的分子信息(参考模块102)。所获得的标度基于遍及感兴趣的分子Mi的所有数据集而确定的最低强度和最高强度。

[0075] 当然,能够产生超过三个的数据集以及对应的图以用于横向地通过对数个数据集J1-Jn的统计组合进行分析,而无需将全部数据集加载到存储器中。然后能够对成像数据进行比较以取出潜在标记。还能够组合具有共同行为的位置以允许数据的新颖解释。

[0076] 换句话说,在单一工具中,本发明对数据进行比较和查询以获得其质量可以计量的更切合的标准化信息。能够使用为使数据标准化而插入到每个图像中的质量控件来加载数个数据集。

[0077] 此外,该过程能够在需要时调整具有不同空间大小的同一生物样本的数据集J1和J2的位置的大小,以便使得该位置与具有最细空间大小的数据集相校准。能够将来自质谱分析仪(成像MALDI)的图处理与来自PET(正电子发射断层扫描)或MRI(磁共振成像)类型的系统的图处理相组合。

[0078] 还能够在单个图像中共同地表示数个数据集J1-Jn。能够进行“为当前图的领域z和a选择所有位置x和y(具有相同的空间大小)的所有值v、w”的类型的请求。领域z和a表示两个不同的成像技术的值。

[0079] 还能够执行能够通过将请求插入到数据库BDD中而添加的临时统计处理。以这种方式,根据本发明的过程提供了到了如下程度的较大灵活性:能够在分析过程的任何时刻添加关于图像的每个位置的信息。

[0080] 该过程基于客户端/服务器架构。能够通过单个接口来查询数据库BDD。在单用户架构中、或者在本地或经由网络(内部的或外部的)可访问的服务器的集群中来存储以及查询数据。客户端接口软件能够被安装在工作站处或者在web接口版本中。本发明的另一目的是数据服务器10,所述数据服务器10包括存储软件指令的存储器,所述软件指令用于执行处理多个数据集J1-Jn的过程的至少一些步骤。

[0081] 并且,数据的已处理数目可能容易地增长,使得待分析的数据量不再对分析构成限制。不管数据库BDD的大小如何,数据的插入都是有效的。

[0082] 当然,已仅通过示例的方式给出了以上描述,并且以上描述不限制本发明的领域,而且以所有其它等同物来代替执行细节将不超过其范围。

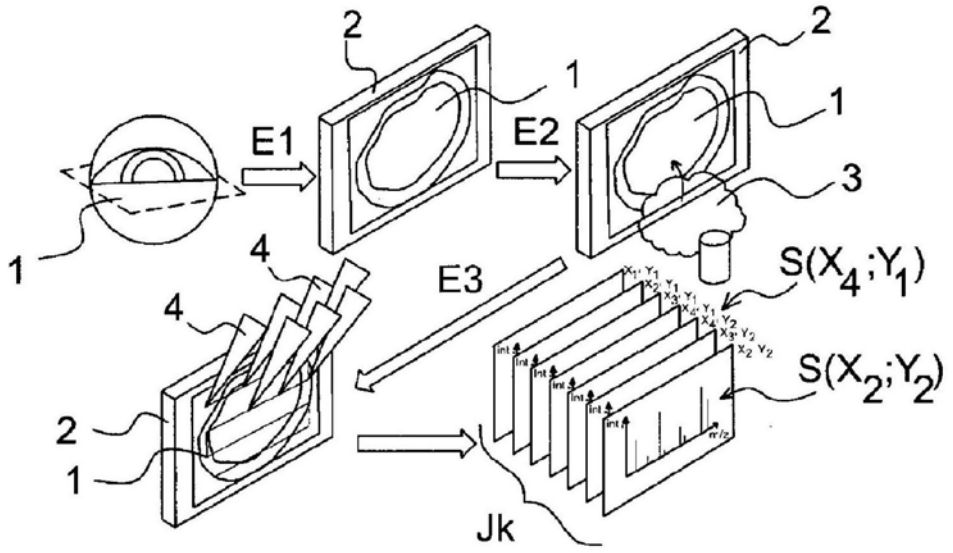

2
1
E1
E2
2
1
3
1
4
4
E3
S(X4;Y1)
S(X2;Y2)
2
1
Jk

图1

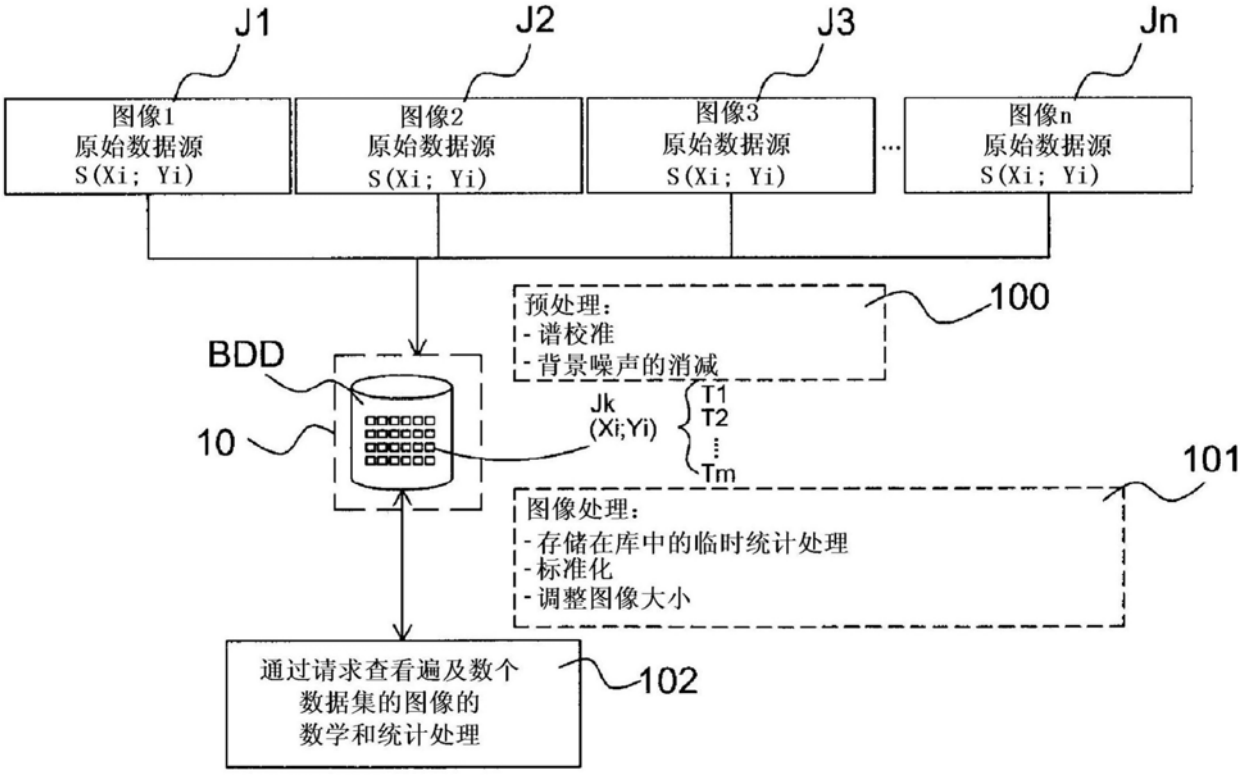

J1
J2
J3
Jn
图像1
原始数据源
S(Xi; Yi)
图像2
原始数据源
S(Xi; Yi)
图像3
原始数据源
S(Xi; Yi)
图像n
原始数据源
S(Xi; Yi)
100
预处理:
- 谱校准
- 背景噪声的消减
BDD
10
Jk
(Xi;Yi)
T1
T2
Tm
101
图像处理:
- 存储在库中的临时统计处理
- 标准化
- 调整图像大小
通过请求查看遍及数个
数据集的图像的
数学和统计处理
102

图2

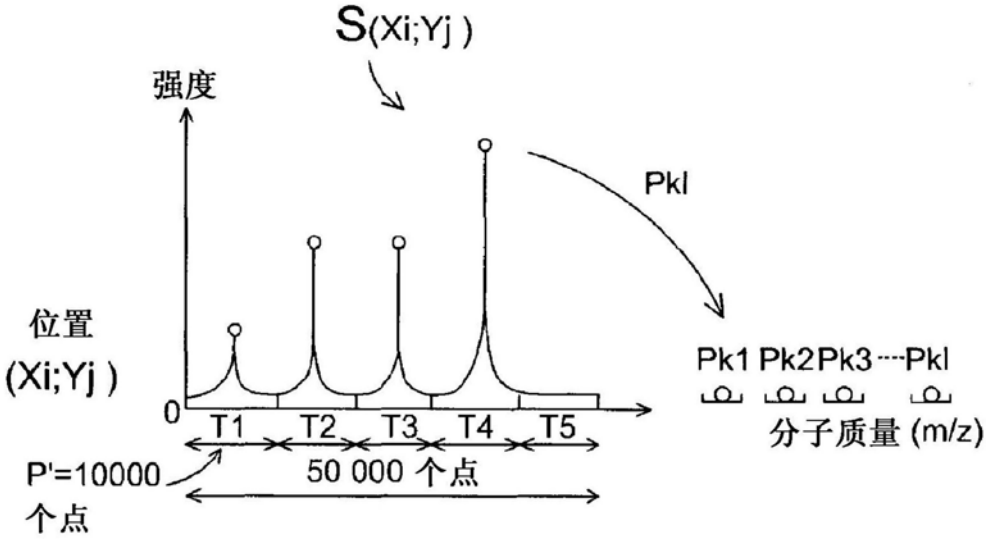

S(Xi;Yj)
强度
Pkl
位置
(Xi;Yj)
0
T1
T2
T3
T4
T5
Pk1 Pk2 Pk3 ···· Pkl
分子质量 (m/z)
P'=10000
个点
50 000 个点

图3

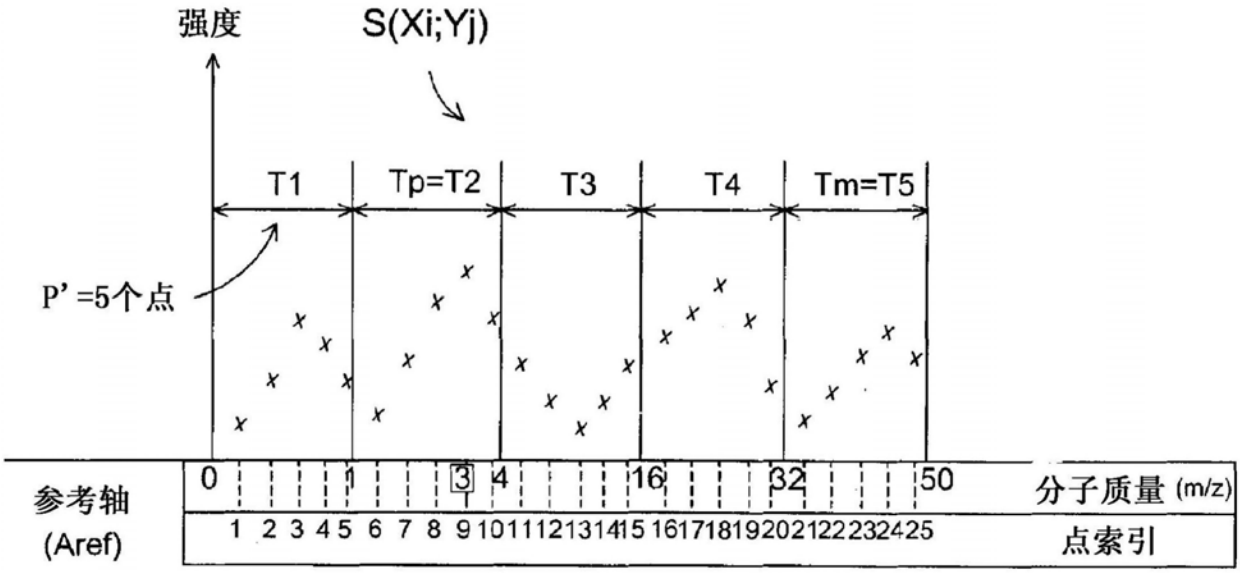

强度
S(Xi;Yj)
T1
Tp=T2
T3
T4
Tm=T5
P' =5个点
参考轴
(Aref)
0
1
3
4
16
32
50
分子质量 (m/z)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
点索引

图4

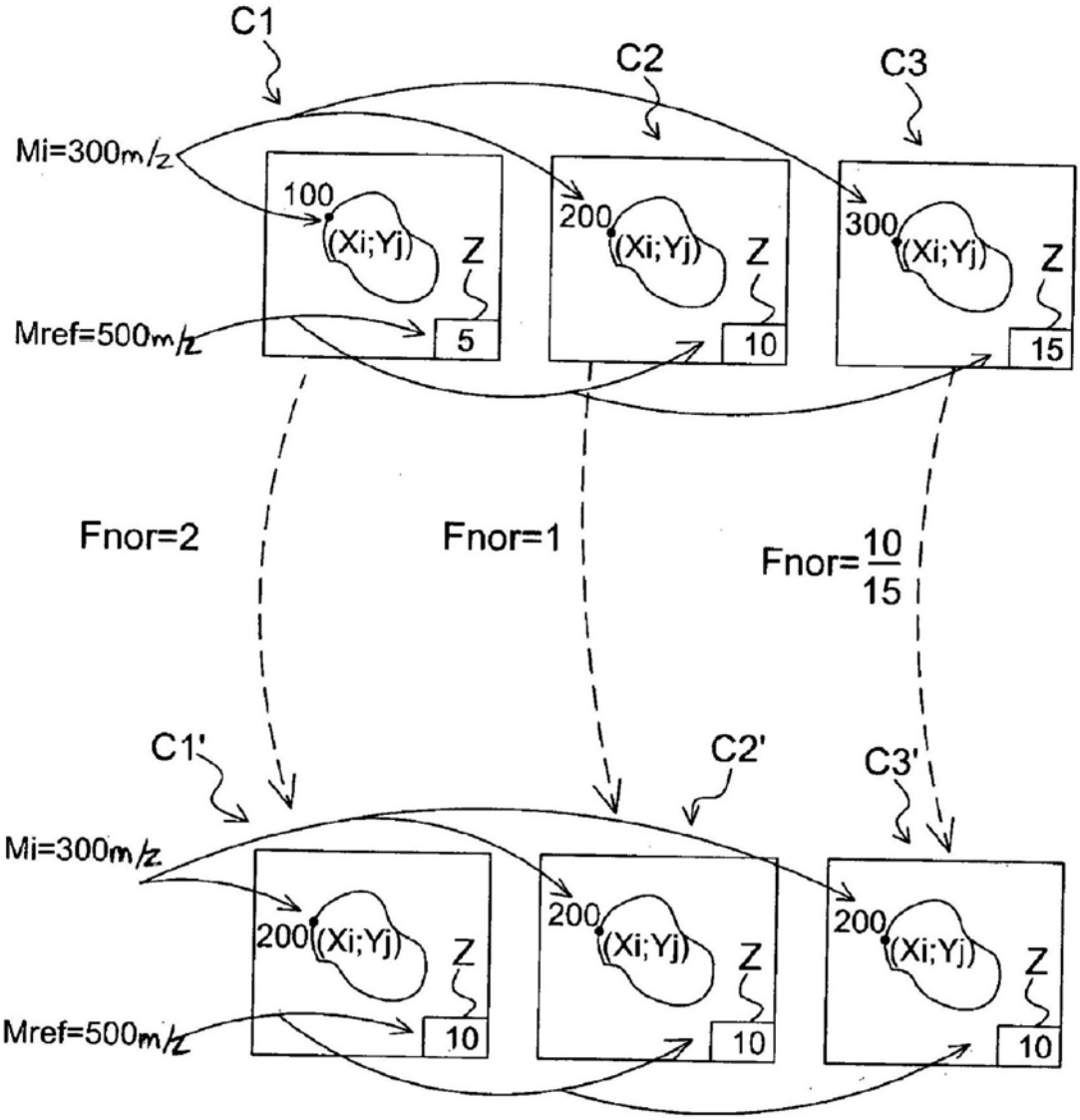

C1
C2
C3
Mi=300m/z
100
(Xi;Yj)
Z
5
200
(Xi;Yj)
Z
10
300
(Xi;Yj)
Z
15
Mref=500m/z
Fnor=2
Fnor=1
Fnor=10/15
C1'
C2'
C3'
Mi=300m/z
200
(Xi;Yj)
Z
10
200
(Xi;Yj)
Z
10
200
(Xi;Yj)
Z
10
Mref=500m/z

图5