



[12] 发明专利说明书

专利号 ZL 00801913.4

[45] 授权公告日 2007 年 4 月 18 日

[11] 授权公告号 CN 1311386C

[22] 申请日 2000.6.29 [21] 申请号 00801913.4

[30] 优先权

[32] 1999. 7. 2 [33] US [31] 09/347,214

[86] 国际申请 PCT/EP2000/006082 2000. 6. 29

[87] 国际公布 WO2001/003002 英 2001. 1. 11

[85] 进入国家阶段日期 2001. 5. 8

[73] 专利权人 皇家飞利浦电子有限公司

地址 荷兰艾恩德霍芬

[72] 发明人 G·韦迪林加姆

M·阿布德尔-莫塔莱布

[56] 参考文献

CN1151798A 1997. 6. 11

CN1137320A 1996. 12. 4

US5806061A 1998. 9. 8

MULTIMEDIA CONTENT DESCRIPTION IN THE INFOPYRAMID CHUNG. SHENG LI ET AL, ACOUSTICS, SPEECH AND SIGNAL PROCESSING, 1998. PROCEEDINGS OF THE 1998 IEEE INTERNATIONAL CONFERENCE ON SEATTLE, WA, USA 12. 15 MAY 1998, NEW YORK, NY, USA, IEEE, US 1998

审查员 刘宇儒

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 吴立明 梁永

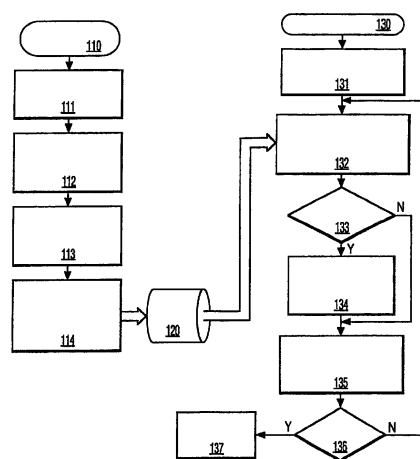
权利要求书 4 页 说明书 15 页 附图 2 页

[54] 发明名称

多媒体信息的元描述符

[57] 摘要

除了描述符之外用元描述符执行多媒体信息检索。“描述符”为特征的表示、“特征”为多媒体信息的区别特征,而“元描述符”为关于描述符的信息。元描述符是通过从多媒体信息中抽取(111)描述符,根据描述符群集(112)多媒体信息,赋予(113)各群集元描述符,及将元描述符附加(114)在注册表中的多媒体信息上,而为注册表中的多媒体信息(10、12、14、16、18、20、22、24)生成的。通过使用仿效实例查询形式的查询(131)获取(132)注册表多媒体项的描述符与元描述符,如果以前未曾生成相同类型的查询描述符则生成(133、134)查询描述符,比较(135)注册表多媒体项与查询多媒体项的描述符,及分级与显示(136、137)结果,来查询多媒体注册表。



1. 一种表示多种多媒体信息（10、12、14、16、18、20、22、24）的方法，包括：

获取多个描述符的步骤（111），所述多个描述符分别是所述多媒体信息的多个特征的表示，所述描述符来自预定的一组不同的描述符类型；

为多媒体信息项的检索，确定所述多个描述符中每一个的相关性；

生成（113）这些描述符的至少一个元描述符，其中所述元描述符包括一组数据元素，每个数据元素指示上述确定的相应描述符的相关性；以及

将该元描述符附加（114）在该多媒体信息上。

2. 权利要求 1 的方法，其中该多媒体信息集体地具有各种内容类型，所述方法进一步包括生成描述符的群集的步骤（112），其中：

该元描述符生成步骤包括生成（113）所述群集的元描述符；以及

该元描述符附加步骤包括分别将群集的元描述符附加（114）到用群集中的描述符所描述的多媒体信息项上。

3. 权利要求 2 的方法，其中该元描述符生成步骤包括为各群集生成指示其中的描述符的相关性的各自的数据元素组。

4. 权利要求 3 的方法，其中：

至少一些描述符为属于多媒体内容类别的多媒体信息项的特征的表示，这些特征构成特征的有序集合，其中包含颜色、结构、形状与草图，并且该多媒体内容类别是静止图象；及

该元描述符生成步骤包括为各群集生成指示其中的描述符的相关性的各自的二进制矢量。

5. 权利要求 3 的方法，其中：

至少一些描述符是属于一种多媒体内容类别的多媒体信息项的特征的表示，这些特征构成特征的有序集合，其中包含颜色、结构、形状与草图，以及该多媒体内容类别是静止图象；及

该元描述符生成步骤包括为各群集分别生成指示其中的描述符的各自的权重的权重值群。

6. 权利要求 3 的方法，其中：

至少一些描述符是属于一种多媒体内容类别的多媒体信息项的特征

的表示, 这些特征构成特征的有序集合, 其中包含颜色、结构、形状与草图, 并且该多媒体内容类别为静止图象; 以及

该元描述符生成步骤包括为各群集生成标识具有预定的表示类型的集合的至少一种相关特征的各自的字符串, 及至少这些相关特征之一来自该预定的表示类型的集合。

7. 权利要求 2 的方法, 还包括将这些描述符指向多媒体信息。

8. 权利要求 1 的方法, 其中:

至少一些描述符是属于一种多媒体内容类别的多媒体信息项的特征的表示, 这些特征构成特征的有序集合, 其中包含颜色、结构、形状与草图, 并且该多媒体内容类别为静止图象; 以及

该元描述符生成步骤包括生成指示其中的描述符的相关性的二进制矢量。

9. 权利要求 1 的方法, 其中:

至少一些描述符是属于一种多媒体内容类别的多媒体信息项的特征的表示, 这些特征构成特征的有序集合, 其中包含颜色、结构、形状与草图, 并且该多媒体内容类别是静止图象; 以及

该元描述符生成步骤包括生成一组指示其中的描述符的权重的权重值。

10. 权利要求 1 的方法, 其中:

至少一些描述符是属于一种类别的多媒体内容的多媒体信息项的特征的表示, 这些特征构成特征的有序集合, 其中包含颜色、结构、形状与草图, 并且该多媒体内容类别是静止图象; 以及

该元描述符生成步骤包括生成标识具有预定的表示类型的集合的至少一种相关特征及来自该预定的表示类型的集合的至少一种表示类型的字符串。

11. 一种利用查询多媒体信息项在使用描述符描述的注册表中搜索多媒体信息 (10、12、14、16、18、20、22、24) 的方法, 包括:

获取注册表描述符的元描述符的步骤 (132), 其中所述元描述符包括一组数据元素, 每个数据元素根据为多媒体信息项的检索的所述描述符的一个确定的相关性来指示各个注册表描述符的相对权重;

选择 (131) 查询多媒体信息;

根据元描述符从查询多媒体信息项中抽取 (134) 至少一个查询描述

符以获得至少一个查询描述符;

将该查询描述符与注册表描述符比较; 以及

按照比较步骤排定(137)注册表中至少一些多媒体信息的次序。

12. 权利要求 11 中的方法, 其中这些注册表描述符是多个不同描述符类型的实例, 及其中:

从查询多媒体信息抽取至少一个查询描述符的步骤包括根据元描述符抽取(133、134)多个描述符以获取多个查询描述符, 这些查询描述符是描述符类型的各自的实例; 以及

该比较步骤包括将查询描述符与具有相同描述符类型的各自的注册表描述符比较(135、136)。

13. 权利要求 11 的方法, 其中所述获取元描述符的步骤包括:

从注册表中的多媒体信息获取注册表描述符的步骤(111);

生成注册表描述符的群集(112);

生成(113)群集的元描述符, 其中所述群集的元描述符包括表示其中注册表描述符的相关性的一组数据元素;

将群集的元描述符附加(114)到群集中各自的多媒体信息上。

14. 权利要求 13 的方法, 其中这些注册表描述符是多个不同描述符的实例, 及其中:

从查询多媒体信息抽取至少一个查询描述符的步骤包括根据元描述符抽取(133、134)多个描述符以获取多个查询描述符, 这些查询描述符是描述符类型的各个的实例; 以及

该比较步骤包括将查询描述符与具有相同描述符类型的各自的注册表描述符比较(135、136)。

15. 一种利用查询多媒体信息项搜索多媒体信息(10、12、14、16、18、20、22、24)的系统, 包括:

一个用于存储多媒体信息的注册表, 所述注册表中的多媒体信息由多个不同类型的注册表描述符描述, 所述注册表描述符是多媒体信息的特征表示, 且来自不同描述符类型的预定组; 和

一个含有注册表描述符的元描述符的存储器, 其中所述元描述符包括一组数据元素, 每个数据元素根据为多媒体信息项的检索的描述符的一个确定的相关性来指示各个注册表描述符的相对权重;

所述系统用于:

根据元描述符从查询多媒体信息项中抽取(134)至少一个查询描述符以获得至少一个查询描述符;

将该查询描述符与注册表描述符比较(135); 以及

按照比较步骤排定(137)注册表中至少一些多媒体信息的次序。

16. 权利要求 15 的系统, 其中:

多媒体内容的类别是静止图象(12);

这些特征是包括颜色、结构、形状与草图的特征有序集合; 以及

这些数据元素是二进制矢量的位。

17. 权利要求 15 的系统, 其中:

多媒体内容的类别是静止图象(12);

这些特征是包括颜色、结构、形状与草图的特征的有序集合; 以及

这些数据元素是权重。

18. 权利要求 15 的系统, 其中:

多媒体内容的类别是静止图象(12);

这些特征是包括颜色、结构、形状与草图的特征的有序集合; 以及

这些数据元素是一串字符。

19. 权利要求 18 的系统, 其中该串包括:

用于标识相关特征的第一字符, 该相关特征具有预定的表示类型的集合;

第二字符, 用于表示利用来自预定的表示类型的集合的多少表示类型来表示该相关特征; 以及

多个第三字符, 用于标识来自该预定的表示类型的集合的表示类型的特定类型。

多媒体信息的元描述符

技术领域

本发明涉及多媒体数据的基于内容的处理,而更具体地涉及建立与使用描述其内容的多媒体数据属性。

背景技术

多媒体信息通常以多种不同形式存在,其中包含诸如数字、模拟(如 VCR 磁带与录音磁带)、光(如传统影片)、图象(如纸上的画与图)、等。在现代社会中定位这种多媒体信息的能力是重要的,而在诸如教育、新闻业(例如利用他的姓名、话音或面貌搜索某一政客的讲话)、旅游信息、文化服务(如历史博物馆、艺术陈列馆等)、娱乐(如搜索比赛或卡拉 OK 名称)、调查工作(如人类特征识别与法医学)、地理信息系统、遥感(如地图绘制学、生态学、自然资源管理等)、监视(如交通控制、地面运输、敌对环境中的非破坏性探测等)、生物医学应用、购物(如搜索你所喜爱的服装)、建筑、房地产、内部设计、社交(如约会服务)从及影片、电视与无线电档案等各种专业与消费者应用中尤为重要。不幸的是,现有的系统在搜索多媒体信息中并不彻底、快速或高效;见诸如国际标准化组织 ISO/IEC JTC1/SC29/WG11 电影与音频编码, MPEG-7 应用文献卷 8, N2728 号, 1999 年 3 月, 其全文在此作为参考。

支持搜索多媒体信息的重要步骤为以采用现代计算机系统能搜索的方式来表示它。在研制超越简单的波形或基于样本的表示、诸如 MPEG-1 与 MPEG-2 等基于压缩的表示、及诸如 MPEG-4 等基于对象的表示,以及能传递到设备或计算机代码或由其访问的音像信息表示形式中已表示了很大关注。已为描述多媒体内容及抽取表示与查询得出的表示的集合开发了许多专利解决方法,但它们只是繁殖了更多的不同多媒体信息并加重了进行快速与高效的多媒体信息搜索的困难。

“描述符”是特征的表示,“特征”是多媒体信息的区别性特征,不论多媒体信息的介质或技术,也不论如何存储、编码、显示与传输该多媒体信息。由于用在不同的专利多媒体信息检索中的描述符不一定兼

容,对建立支持制作、交换、检索与/或重复使用多媒体信息的计算系统的操作要求的描述多媒体内容数据的标准表示了关注。实例中包含为图象理解(如监视、智能视觉、智能照相机)、媒体转换(如语音到正文、画面到语音、语音到画面)、与信息检索(用户感兴趣的各種类型的多媒体文献的快速与高效搜索)以及音像内容描述流中的过滤(只接收满足用户的偏爱的多媒体数据项目)设计的计算系统。

从而,存在着对支持这些操作要求以及尚待开发的其它操作要求的描述多媒体内容数据的标准的需求。

发明内容

从而,本发明在具体实施例中实现的目的是改进从注册表中检索多媒体信息的效率。

本发明在具体实施例中实现的另一目的为改进从注册表中检索多媒体信息的速度。

本发明在具体实施例中实现的又另一目的为提供多媒体信息的特征的标准表示。

本发明提供了一种表示多种多媒体信息的方法,包括:获取多个描述符,所述多个描述符分别是所述多媒体信息的多个特征的表示,所述描述符来自预定的一组不同的描述符类型;为多媒体信息项的检索,确定所述多个描述符中每一个的相关性;生成这些描述符的至少一个元描述符,其中所述元描述符包括一组数据元素,每个数据元素指示上述确定的相应描述符的相关性;以及将该元描述符附加在该多媒体信息上。

本发明还提供了一种利用查询多媒体信息项在使用描述符描述的注册表中搜索多媒体信息的方法,包括:获取注册表描述符的元描述符,其中所述元描述符包括一组数据元素,每个数据元素根据为多媒体信息项的检索的所述描述符的一个确定的相关性来指示各个注册表描述符的相对权重;选择查询多媒体信息;根据元描述符从查询多媒体信息项中抽取至少一个查询描述符以获得至少一个查询描述符;将该查询描述符与注册表描述符比较;以及按照比较步骤排定注册表中至少一些多媒体信息的次序。

本发明还提供了一种利用查询多媒体信息项搜索多媒体信息的系统,包括:一个用于存储多媒体信息的注册表,所述注册表中的多媒体信息由多个不同类型的注册表描述符描述,所述描述符是多媒体信息的

特征的表示，且来自不同描述符类型的预定组；和一个含有注册表描述符的元描述符的存储器，其中所述元描述符包括一组数据元素，每个数据元素根据为多媒体信息项的检索的所述描述符的一个确定的相关性来指示各个注册表描述符的相对权重；所述系统用于：根据元描述符从查询多媒体信息项中抽取至少一个查询描述符以获得至少一个查询描述符；将该查询描述符与注册表描述符比较；以及按照比较步骤排定注册表中至少一些多媒体信息的次序。

在本发明的各实施例中实现这些与其它目的。

附图说明

图 1 为注册表中的多媒体数据的各种形式的图形表示。

图 2 为按照本发明的元描述符生成进程与多媒体查询进程的流程

图。

图 3 为按照本发明的示例关系数据库的记录表。

图 4 为按照本发明的改进注册表中的多媒体元描述符的进程的流程图。

具体实施方式

图 1 中示出多媒体数据可存在的各种形式及包含例如图形 10、静止图象 12、视频 14、影片 16、音乐 18、讲演 20、声音 22 与其它媒体 24 的多个出现的示例。这多个出现了不同地存储，不同地编码，不同地传输，存在在不同的介质上，或者用不同技术产生的。多媒体数据可存储在一处或分布在全球上；例如，用户所关注的数字化多媒体可存储在自含的关系或面向对象数据库中，或在以不同技术实现并存储在分散在全球并只能在因特网上才能访问的不同专用计算机的独立的数据库中。当然，用户所关注的非数字多媒体可类似地存储在单一实体控制下的一个集合中，或广泛地散布在不同实体控制下的不同集合中。不论多媒体信息的存在形式及它是如何存储的，为了高效地搜索特定的多媒体数据，用户会希望将多媒体信息的集合看作单一的注册表，如参照数字 1 所示。

我们已发现除了描述符之外还利用元描述符的多媒体信息检索不仅在标识多媒体信息中高效并且还能标识以多种多样不同方式表示的多媒体信息。“描述符”是特征的表示，“特征”是多媒体信息的区别性特征，而“元描述符”是关于描述符的信息。元描述符不同于元数据但与

元数据的一般概念相关，它是嵌入附加信息的著名方法。例如，文献中的元数据可包含文献中的图象的格式，而数据库中的元数据可包含关系中的属性的值限制或统计信息。具体地，多媒体信息项的元描述符标识包含用于标识多媒体信息项的最有用的信息的该多媒体信息项的描述符的部分。元描述符的概念是基于给定的多媒体信息项最有资格知道用什么来描述它最好，及指示这一信息极大地增进基于内容的检索的前提的。有利地，由于元描述符的大小通常较小，元描述符使计算机化搜索多媒体信息能更快地完成，并且由于消除了不大相关的信息而更高效。

虽然许多多媒体检索技术适用于采用元描述符，采用元描述符搜索多媒体信息的较佳技术为情况搜索或仿效实例查询。在示例性图象的仿效实例查询中，指定特定的静止图象作为查询该注册表的基础。查询是由作为人类用户或自动化过程的查询的发动者指定的，或由在检索进程中使用的检索算法指定的。查询是根据在注册表中的多媒体信息的元描述符中指示为重要的一或多种特征构成的。一或多种特征的描述符是从查询多媒体信息中抽取的并与从注册表多媒体信息中的描述符比较以获得类似性测度，将它们用于选择一或多个“匹配的”多媒体信息项。在一些实例中，注册表多媒体信息的元描述符可立即指示明显的及重大的内容不同性，借此避免执行描述符抽取及查询与特定注册表多媒体信息的特征比较的计算。如果用户不是信息检索专家，为了使用简单起见，用在匹配的多媒体信息检索中的一或多种特定特征最好是对用户透明的。

图 2 为展示用于生成注册表多媒体信息的元描述符及用于执行注册表查询的进程的流程图。虽然元描述符可用于任何类型的多媒体信息，图 2 中所示的实例是基于静止图象的以便说明。用基本步骤 111-114 说明生成元描述符的方法 110，并用基本步骤 131-137 说明图 1 中所示的多媒体注册表的查询方法 130。虽然元描述符也可通过由人或半自动的技术的混合物形式化用户输入来生成，元描述符生成方法 110 为机器学习的不受监视的或自动化的方法。查询方法 130 除了查询的用户阐述之外最好是自动化的。描述符生成与多媒体信息检索的各方面是众所周知的并在下述各种出版物中描述，其中包含例如 Yong Rui、Thomas S.Huang 与 Shih-Fu Chang 的“图象检索：过去、现在与未来”，可视通信与图象表示期刊，10，1-23（1999）；Sharad Mehrotra、Yong Rui、Michael

Ortega-Binderberger 与 Thrmass S.Huang 的“在 MARS 中的图象上支持基于内容的查询”，IEEE 多媒体计算与系统国际会议会报，1997 年 6 月 3-6 日，Chateau Laurier,Ottawa,Ontario,Canada,1997,632-633 页；Sharad Mehrotra、Yong Rui、Kaushik Chakrabarti、Michael Ortega-Binderberger 与 Thomas S. Huang 的“多媒体分析与检索系统”，第三次信息检索系统国际讨论会会报，Como,Italy,1997 年 9 月 25-27 日,39-45 页；及 Patrick M. Kelly、Michael Cannon 与 Donald R. Hush 的“用图象实例的查询：CANDID 方法”，图象与视频数据库Ⅲ的存储与检索，SPIE 卷 2420，1995，238-247 页，这时通过引用将它们全文结合在此。

元描述符生成方法 110 中的第一步骤 111 为从多媒体（“MM”）信息中抽取描述符。描述符抽取算法与软件以及索引算法与软件是本技术中众所周知的，并在上面的出版物中描述了它们的实例。除外任何相关标准施加的限制，可根据需要采用任何一种描述符抽取技术或其组合。通常，从静止图象抽取的描述符例如包含颜色、形状、结构与草图，但在任何一种实例中抽取的特定描述符取决于在自动化过程中所采用的软件，用户在手动过程的动作，或在混合过程中用户与软件的交互作用。自动化过程是有利的因为它们能处理注册表中的大量多媒体信息而无须太多的人类干预。通常在静止图象上操作的自动化过程并不在语义级上操作，即它们并不在这些项中描述系列画像，而是生成整个静止图象的或将该图象分成的多个块的颜色、形状、结构与草图的值。

下一步骤 112 为根据描述符将多媒体信息群集在一起，虽然如果愿意也可采用其它技术，在上面提到的出版物中描述了其实例。主要，群集是根据施加在描述符上的一定标准将来自大的混合数据集的类似的多媒体信息组合，对于小的相同内容类型信息的注册表不需要群集。一个群集是相似实体的集合，而来自不同群集的实体是不相似的。可以用受控制或不受控制的群集完成元描述符的抽取。给定早先已赋予标记或描述符的群集代表的集合，用受控制的群集抽取元描述符包含根据其特征群集多媒体内容。然后赋予各群集中的图象该群集的代表标记或描述符。用不受控制的群集抽取元描述符包含根据各描述的特征群集多媒体内容。例如，一个图象集合具有根据它们的颜色特征的第一群集表示及具有根据它们的结构特征的第二群集表示。例如，根据用编程的尺度的各特征的群集的比较来计算群集是如何清楚地定义与小巧的，找出一种

或少数特征在描述给定的图象中胜过其它特征。例如，某一图象根据颜色特征可能属于群集组中非常小巧与清楚地定义的群集，但根据结构特征可能属于群集组中的广泛分散与重叠的群集。从而选择颜色特征作为该图象的元描述符，由于它在它们各自的特征空间中比结构特征更好地分类该图象。

数学上，群集是测试空间中的点的聚集，使群集中任何两点之间的距离小于该群集中的任何一点与不在其中的任何一点之间的距离。见诸如 Anil K. Jain 与 Richard C. Dubes 的“群集数据算法” Prentice Hall 高级参考书系列，1988 1 页。群集分析是将对象分成在特定问题的角度上有意义的子集的过程。借此将这些对象组织成刻画所抽样的总体的特征的高效表示。对象之间的关系是以接近度矩阵表示的，其中行与列对应于对象。如果将这些对象表征为模式、或 d 维量度空间中的点，接近度可以是点对之间的距离，诸如欧氏距离。通常，接近度矩阵是对群集算法的唯一输入。群集在一起的对象可能是不同物种的植物，数字图象中的象素或不同主题的文档。群集分析在模式识别与图象处理中找到若干应用。群集分析在包含受控制的学习的应用中及在人工智能中的辅助机器学习与知识表示中用于特征选择。

从静止图象抽取的描述符通常是表示空间中的点的集合的多维数字的矢量。在简单的情况中，群集算法比较 2D 空间中的点的集合中的距离来确定这些点的接近程度。在更高级的算法中，将这一概念扩展到多维空间。为了保证相容性，最好只将群集过程作用在用同一抽取算法处理的多媒体信息上。对于静止图象，群集是基于通常低级特征的相似性的；例如，可将呈现特定颜色与结构的清楚的模式（例如诸如从海滩风景生成的）而带有草图与形状的有关不定结果的某些图象认为是在同一群集中。不受控制的群集算法通常交互式工作，求精它们的结果直到达到用户规定的阈值点为止。

下一步骤 113 为将元描述符赋予各群集。例如，当只为颜色形成模式时，将 100% 颜色的元描述符赋予群集。当只为草图形成模式时，这可预期用于单色工程制图，例如，可将 100% 草图的元描述符赋予群集。对于前面的例子中只为特定的颜色与结构形成清楚的模式静止图象的特定群集，则将 50% 颜色与 50% 结构的元描述符赋予群集。如果希望过程自动化，可用根据来自带多媒体信息的过去的经验的统计信息的直观

规则赋予权重。作为替代,如果愿意,可以手动或与人交互作用的半自动方式赋予元描述符。

元描述符可采用对程序员方便的任何形式。在一种特别紧凑的形式中,元描述符为二进制矢量 X , 给定了该多媒体内容类别的固定数目的有序的特征, 各位 X_i 表示特征的相关性 (X_i 第 i 特征相关的)。在静止图象的情况下, 适用的矢量为四位矢量, 其中二进制 1 或 0 分别表示描述多媒体信息的颜色、形状、结构与草图的重要性或无关性。如果在特定静止图象中只有颜色是重要的, 适当的元描述符为 1000。

能赋予特定权重的一种符号利用加权矢量 X , 其中给定了该类别的多媒体内容的固定数目的有序的特征, 矢量的各元素 X_i 表示赋予第 i 特征的权重。如果说颜色与草图都重要但具有不同的权重, 这一类型的适当元描述符为 “70, 0, 0, 30”, 表示颜色具有 70% 权重而草图具有 30% 权重。

另一种形式是串符号, 它不仅能处理不同权重也能处理不同抽取算法与将静止图象分成多块。此外, 串表示允许以后在元描述符中考虑新特征。在串符号中, 给定按照预定义的标准符号的一组特征, 串中的各字符或字符群表示特征的相关性。例如, 考虑四种有效颜色描述符类型的组: (1) RGB 颜色空间中整个图象的单一直方图; (2) 分成 5×5 格的 RGB 颜色空间中的图象的 25 个直方图, 所得出的 25 个块中的各个用一个直方图表示; (3) YUV 彩色空间中的整个图象的直方图; 以及 (4) 分成 5×5 格的 YUV 彩色空间中的图象的 25 个直方图, 所得出的 25 块中的每一个用一个直方图表示。假定这些描述符类型是从 1 至 “ n ” 用数字排序的, n 为有效描述符类型的数目, 这里为 4。例如用第一与第四颜色描述符最佳描述的静止图象的适当的串元描述符为 “C214”, 其含义为: 颜色特征是相关的 (C) 并且是从预定义的颜色表示的组中用两种 (2) 颜色得出的, 即来自预定义的颜色表示组中的第一与第四 (14) 颜色表示。串符号特别灵活, 不仅允许标识不同的彩色空间 (例如 RGB 与 YUV) 并且还允许不同地计算各彩色空间 (例如, 作为一块、10 块的组、100 块的组、等)。串符号的扩展还能用适当的预定义代码处理不同的抽取算法。

不同形式的元描述符可用于不同的多媒体信息, 而在特定形式的元描述符中未提供的任何信息能用默认提供。例如, 如果采用默认描述符

抽取方法与默认彩色空间，二进制元描述符已经足够。

下一步骤 114 是根据群集信息将元描述符附加在多媒体信息上。各种不同的“附加”技术是众所周知的并可根据媒体类型与访问它的方式选用，并且在特定的元描述符数据库中可使用不同的附加类型。描述符本身可以存在或不存在，然而如果它们不存在，系统必须知道如何用默认或该元描述符符号中的值计算它们。最好至少将元描述符与它们的附加数据存储在存储器 120 中，它可以是系统能访问的任何类型的数据库。可将描述符存储在存储器 120 中或与从其抽取它们的多媒体信息一起存储。多媒体信息本身位于注册表中（图 1），它可以与存储器 120 中的其它存储空间一样明确或者与因特网一样分散，或甚至分散到包含诸如纸等非电子形式的存储物。

一旦选择了要出现在诸如图象等的元描述符中的特征，如果愿意，元描述符可以允许存在来自预定义的集合中的其它特征。为了减少要测试的特征数，适当时可使用从加标记的训练集合中导出的一组结合规则。例如，特定的注册表可包含不是用草图很好地描述的多媒体信息，因此检索系统不需要使用草图。

抽取期间，如果愿意可组合来自多媒体内容的特征，用较高级语义信息来确定元描述符中的值。例如，当多媒体内容为图象时，将来自图象的特征与来自与该图象关联的文字说明的语义信息组合来确定该图象的元描述符中的值。

图 3 为使用各种附加技术的关系数据库文件（任何其它类型的数据库同样适用）的简单多媒体信息表。示例性地，该表具有五个字段，多媒体信息号字段 **MM_INFO_NO**、描述符值字段 **DV**、元描述符值字段 **MDV**、多媒体文件源字段 **MM_SOURCE**、及注释字段 **COMMENT**。**MM_INFO_NO** 字段为主键字段。**DV** 与 **MDV** 字段为用于包含例如串矢量的字符字段。**MM_SOURCE** 字段为链接到或嵌入诸如数字化文档、图、画面、声音等 OLE 对象的 OLE 数据类型。**COMMENT** 字段为摘要数据类型字段。假定将描述符嵌入多媒体信息中或检索系统能从多媒体信息中抽取描述符，如果愿意可从表中省略 **DV** 字段。然而，在本地数据库中具有描述符将使检索系统能更快地操作。

用主键 **MMO1** 标识的记录包含描述符值 **DV01** 与元描述符值 **MDV01**，并用 OLE 数据类型字段中的链路附加在存储在 **FILE01** 中的

数字化的静止图象上。用主键 MM02 标识的记录包含元描述符值 MDV02, 并附加在用 OLE 数据类型字段中的链路附加在存储在 FILE02 中的数字化静止图象上。描述符值是在查询注册表中的多媒体信息的过程中从 FILE02 的内容中抽取的。用主键 MM03 标识的记录包含图象中的一块多媒体信息的描述符值 DV03 与元描述符值 MDV03, 并用 OLE 数据类型字段中的链路附加在存储在 FILE03 中的数字化静止图象上。用主键 MM04 标识的记录包含同一图象中的另一块多媒体信息的描述符值 DV04 与元描述符值 MDV04, 并用 OLE 数据类型字段中的链路附加到存储在 FILE03 中的数字化静止图象上。用主键 MM05 标识的记录以 OLE 数据类型字段中的链路附加到存储在 FILE05 中的数字化静止图象上, 这一记录的 DV 与 MDV 字段是空的, 由于 DV 与 MDV 嵌入在链接文件中并能容易地从其读出。用主键 MM06 标识的记录以 OLE 数据类型字段中的 URL 链路附加到存储在可在因特网上访问的另一数据库中的数字化静止图象上。这一记录的 DV 与 MDV 字段是空的, 由于 DV 与 MDV 嵌入在链接的文件中并能从其读出。用主键 MM07 标识的记录包含元描述符值 MDV07, 并连接在 VCR 磁带上, 这一记录的 DV 字段是空的, 由于该描述符值嵌入在 VCR 磁带的垂直消隐间隔中并能从其读出, 这一记录的 MM_SOURCE 字段是空的。除非检索系统从元描述符检测到在查询多媒体与 VCR 磁带之间存在内容上的明显与重大的不同, 必须安装该磁带并必须在查询期间从该 VCR 磁带读出描述符。用主键 MM08 标识的记录包含描述符值 DV08 与元描述符值 MDV08, 并连接在印在照相纸上及存档在抽屉 08 中的静止图象上。这一记录的 MM_SOURCE 字段是空的。

用基本步骤 131_137 示出诸如图 1 中所示的多媒体注册表的查询方法 130。步骤 131 为用户用任何传统方法(这里是仿效实例查询)构成查询。在仿效实例查询中, 用户选择一多媒体信息项并希望从注册表中找出所有匹配的多媒体信息项。在步骤 132 中, 示例性地从存储器 120 中检索注册表中的一个多媒体信息项的描述符与元描述符。在描述符的情况中, 它们可替代地与该多媒体信息一起存储并因此必须从该多媒体信息检索, 或者可能得不到它们而必须根据元描述符中的值重新抽取。如果注册表多媒体信息项的描述符具有未曾早先在查询 130 中处理过的类型(步骤 133 为是), 便通过应用抽取方法与由该注册表多媒体信息项的

元描述符所指示的权重从该查询多媒体项中抽取对应的描述符（步骤 134）。然后在查询描述符与注册表多媒体信息项描述符之间进行比较（步骤 135）。在注册表多媒体信息项的元描述符中未给出权重的特征无须为元描述符抽取处理。对数据库中所有群集重复这一比较（步骤 136 为否），并用本技术中众所周知的适当装置适当地分级来自各群集的最接近的匹配的集合，并将其显示给用户（步骤 137）。

图 4 中示出包含人类专家的形式化用户输入的优化元描述符的技术。为了描述清楚起见，假设简单的静止图象注册表。这些图象通常可用少数类别分类，例如人物、植物、风景与纺织品，并允许用诸如颜色、形状、组织与草图等少数方法的基于内容的检索。数据库中给定的图象最好用一或多种这些特征描述，而用其它特征描述则不良。例如人物最好用形状描述，植物最好用颜色与结构描述，而风景最好只用结构描述。用下述人类输入训练具有这些特征的数据库。在步骤 301 中，利用例如颜色、形状、结构与草图等所有特征执行元描述符生成进程，图 2 的步骤 112 - 114 说明这一进程。在步骤 302 中，用例如颜色、形状、结构与草图等所有特征执行多媒体查询进程，图 2 的步骤 131 - 136 说明这一进程。在步骤 303 中，检索系统分级这些结果并将其显示给人类专家。从最接近的匹配中，用户确定哪种方法最适合查询图象，或如果一种以上方法适合查询图象，用户用适当的标准确定各适用特征的权重。用户还指示要给予该适当特征的相同权重的最接近的匹配的集合中的所有其它图象。在步骤 306 中，检索系统根据人类专家的输入例如通过赋予特征新的权重更新元描述符的值。可以采用各种迭代学习技术中任何一种。在训练中完全没有考虑过的图象，可赋予其所有特征相等的权重。重复步骤 301、302、303 与 306 直到人类专家对结果满意为止，此时优化了元描述符而进程 300 结束。

虽然特定的标准化方案不是关键性的，当将元描述符包含进描述符、描述符方案与描述符定义语言的标准的系统中时，元描述符是最有效的。描述符是什么及它在多媒体信息检索系统中如何起作用的说明性观点在下述文件中提出：国际标准化组织 ISO/IEC JTC1/SC29/WG11 动画与声音的编码，MEPG - 7 技术要求文件卷 8，N2727 号，1999 年 3 月，及国际标准化组织 ISO/IEC JTC1/SC29/WG11 动画与声音编码，MEPG - 7；上下文、目的与技术路线图，V.11、N2729 号，1999 年 3

月，通过引用将它们的全文结合在此。按照这些文件中表达的观点，多媒体信息的搜索是通过比较“描述符”与它们的例示（“描述符值”）执行的，描述符是多媒体信息的“特征”的表示而特征为向某人表明某一事物的多媒体信息的区别性特征。描述符定义特征表示的语法与语义。如果愿意，在需要访问不同的相关要求时，可用若干描述符来表示单一特征。例如，颜色特征的可能描述符有：彩色直方图、频率分量的平均值、运动范围、名称的正文、等。通过“描述方案”的机制组合描述符值来构成“描述”。具体地，描述方案（“DS”）规定其分量之间的关联的结构与语义，其分量可以是描述符与描述方案两者，而描述包含 DS（结构）与描述该多媒体数据的描述符值（例示）的集合。描述定义语言（“DDL”）为允许建立新的描述方案与有可能描述符的语言。它还允许扩展与修改现有的描述方案。取自上面提到的 MPEG-7 技术要求文件 V.8 的表 1(修改成包含草图特征)示例特征与其描述符之间的区别。

表 1

特征		描述符	
特征类别类型	特征	特征标记	数据类型
空间	对象纹理, 对象形状		一组多边形顶点
时间	对象的轨迹		链码
客观	对象的颜色		彩色直方图
	对象的形状		一组多边形顶点 一组动差
	对象的结构		一组小波系数, 对比、粗糙度与方向性的集合
	对象的草图		边组
	音频内容		频率分量平均值
主观	感情(喜、怒、愁)		一组本征面参数, 正文
	风格		正文
	注释		正文
制作	作者		正文
	出品人		正文

	导演		正文
组合信息	布景组织		树型图
概念	事件		正文
	活动		正文, 数字值

上面提到的 MPEG-7 技术要求文件中提出的描述符与描述方案的通用要求是受适当的描述符及它们的元描述符支持的。多种类型特征 - 适当的描述符与它们的元描述符用下述各种特征支持多媒体描述, 诸如: N 维时空结构 (如音乐小节的持续时间)、客观特征 (如酒店中的床位数、对象的颜色、对象的形状、音调、等)、主观特征 (如某人如何好、快乐或胖, 主题、风格、等)、制作特征 (如诸如采集的日期、制作者、导演、表演者、角色、制作公司、制作历史、任何非 IPMP 制作信息等关于文件建立的信息)、组织信息 (如如何组成布景、编辑信息、用户的偏爱、等)、及概念 (如事件、活动)。多媒体资料的抽象级 - 支持在不同的抽象级上描述多媒体文献的分层机制, 它提供用户对不同抽象级上的信息的需求, 例如从子对象组成对象, 逐个序列分析视频中的运动, 及视频的曲线结构。交叉模态 - 支持允许根据视觉描述来检索音频数据及反过来的查询的音、像或其它描述符及它们的元描述符 (例如, 当查询为 Pavarotti 的声音的摘录而检索的结果是 Pavarotti 正在唱歌及 Pavarotti 出场的电视剪辑)。多种描述-支持同一资料在其制作过程的若干阶段上的多种描述的处理能力, 以及应用在同一资料的多份拷贝上的描述。描述方案关联 - 适当的描述方案表达描述符与它们的元描述符之间的关联以便能将它们用在一种以上的描述方案中。支持以不同的描述方案编码描述符与它们的元描述符之间的等效关联的能力。描述符优先级 - 描述方案最好支持描述符与它们的元描述符的优先级化以便可以更高效地处理查询。优先级可反映信任度或可靠性级。描述符分层结构 - 适当的描述方案支持不同描述符与它们的元描述符的分层表示以便在 N 级描述符补充 (N-1) 级描述符时可在接连的级中更高效地处理查询。描述符可量测性 - 适当的描述方案支持用它们的元描述符可量测的描述符以便在接连的描述层中更高效地处理查询。时间范围的描述 - 支持不同的时间范围的描述符与它们的元描述符的关联, 分层地 (在全体数据或其时间子集上将描述符与它们的元描述符关联) 以及顺序地

(在相继的时段上将描述符与它们的元描述符接连地关联)两者。直接数据操作 - 支持作为直接引用数据的柄工作的描述符与它们的元描述符,以便能操作多媒体资料。基于正文的描述语言 - 适当的描述符与它们的元描述符指定用在描述中的语言并支持所有自然语言。正文描述中的翻译 - 适当的正文描述提供包含翻译成若干不同语言的方法,以便传递不同语言中的描述之间的关系。

适当的描述符与它们的元描述符支持上面提到的 MPEG-7 技术要求文件中所提出的对描述符与描述方案的功能要求。检索有效性-支持多媒体资料的有效检索。检索效率 - 支持多媒体资料的高效检索。基于相似性的检索 - 支持允许用与查询的相似性程度排列数据库内容次序的描述。关联的信息-支持其它信息与数据的关联。流送与存储的描述 - 支持流送的(与内容同步)及非流送的数据描述两者。分布式多媒体数据库 - 支持分布式数据库中的多媒体数据的同时与透明检索。参照模拟数据 - 支持参数与描述模拟格式的多媒体文献的能力(例如,提供对 VHS 磁带中的序列的时间参照)。交互式查询 - 支持允许交互式查询的机制。链接 - 支持允许空间与时间上定位源数据的机制,包含对相关信息的链路。相关的信息的优先级化 - 支持在上面链接条下提到的允许相关信息的优先级化的机制。浏览 - 支持为了协助用户克服对信息的结构与/或类型的不熟悉性或弄清楚他们的不确定的需要而允许预检信息内容的描述。相关关系 - 支持描述的分量之间的关系。交互性支持 - 支持允许规定关于描述的交互性的措施(例如,关于广播事件的电信投票)。知识产权信息 - 支持包含关于 D、DS 与描述的版权、许可证与证明信息。

适当的描述符与它们的元描述符支持上面提到的 MPEG-7 技术要求文件中提出的描述符与描述方案的视觉特定的要求。特征类型 - 支持允许下述特征(主要涉及查询中使用的信息类型)的视觉描述:颜色、视觉对象、结构、草图、形状、静止与移动图象(如,略图)、容积、空间关系、运动、变形、视觉对象源及其特征(如源对象、源事件、源属性、事件、事件属性、及典型的相关场景与模型(如, MPEG-4 SNHC))。使用描述的数据形象化 - 支持带有提高的形象化方面的能力的多媒体数据描述范围(允许索引的数据的或多或少的粗略的形象化)。视觉数据格式 - 支持下述视觉数据格式的描述:数字电视与影片,诸如 MPEG-1, MPEG-2 或 MPEG-4; 模拟电视与影片; 诸如 JPEG、

寻呼机或其它格式的电子产品中的静止画面；图形，诸如 CAD、3D 模型、特别是 VRML；以及与视频相关的组合数据。尚待定义的其它视觉数据格式的描述也有可能。视觉数据类 - 支持专门应用于下述数据类的描述：自然视频、静止画面、图形、动画（2-D）、三维模型与组合信息。

适当的描述符与它们的元描述符支持上面的 MPEG-7 技术要求文件中提出的描述符与描述方案的音频特定的要求。特征类型 - 音频描述允许下述特征（主要关于查询中使用的信息类型）：频率轮廓（总趋势、旋律轮廓）、音频对象、音色、和声、频率分布、幅度包络、时间结构（包含节奏）、结构内容（典型地语音或歌词）声音逼近（例如通过哼旋律或轰鸣声音效果发出声音梗概的声音）、样品声音（典型的用实例查询）、空间结构（可用于多声道源、立体声、5.1-声道、及各具有特定映象的立体声声音）、声源及其特征（如源对象、源事件、源属性、事件、事件属性、与典型的相关场景）、以及模型（如 MPEG-4 SAOL）。使用描述的数据声音化 - 支持带有增进的声音化方面的能力的多媒体数据描述范围。听觉数据格式 - 支持下述听觉数据类型的描述：数据音频（如 MPEG-1 音频、密致盘）、模拟音频（如乙烯基唱片、磁带介质）、包含通用 MIDI 与卡拉 OK 格式的 MIDI、基于模型的音频（如 MPEG-4 的结构化音频管弦乐队语言 - SAOL）、以及制作数据。听觉数据类 - 支持专用于下述听觉数据的子类的描述：声带（自然音频场景）、音乐、原子声音效果（如拍打）、语音、符号音频表示（MIDI、SNHC 音频）、以及混合信息（包含效果）。

适当的描述符与它们的元描述符支持上面提到的 MPEG-7 要求文件中提出的描述符与描述方案的编码要求。描述高效表示 - 支持数据描述的高效表示。描述抽取 - 元描述符支持按照若干广泛使用的格式的从未压缩与压缩数据容易地抽取的描述符与描述方案的使用。对信息错误与丢失的坚实性 - 支持在传输错误的情况中保证系统的优美的表现的机制。

虽然在上面提到的 MPEG-7 技术要求文件中未提出描述符与描述方案的正文特定的要求，假设正文描述与界面允许查询根据音像描述检索正文数据及反过来，以及对于只有正文的文件的正文描述与包含正文的组合文件的正文描述是相同的，则适当的描述符与它们的元描述符支

持多媒体内容包含或参照除了音像信息之外的正文的能力。

虽然在使用元描述符的某些情况中，搜索引擎或过滤器代理（用户方）可能必须知道元描述生成进程所采用的精确特征抽取算法，在其它情况中在描述生成进程中用于特征抽取的特定算法是与无描述生成进程无关的。因此，元描述进程能提供特征抽取中的技术发展并鼓励竞争性开发。

这时提出的本发明及其应用的描述是示例性的而非旨的限制下面的权利要求中所提出的发明范围。这里公开的实施例的改变与修正是可能的，并且对于熟悉本技术的人员，这些实施例的各种元素的实际替代品与等效物是已知的。可以作出这里公开的实施例的这些与其它改变与修正而不脱离下面的权利要求中所提出的发明范围与精神。

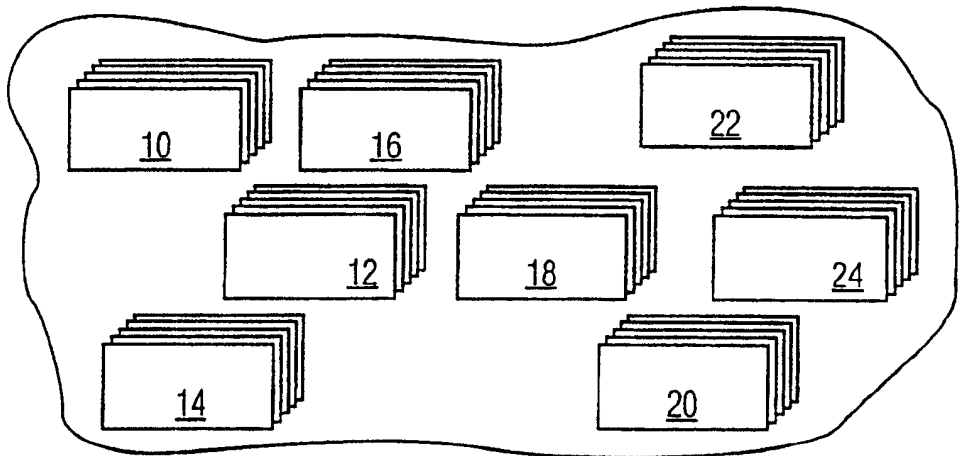


图 1

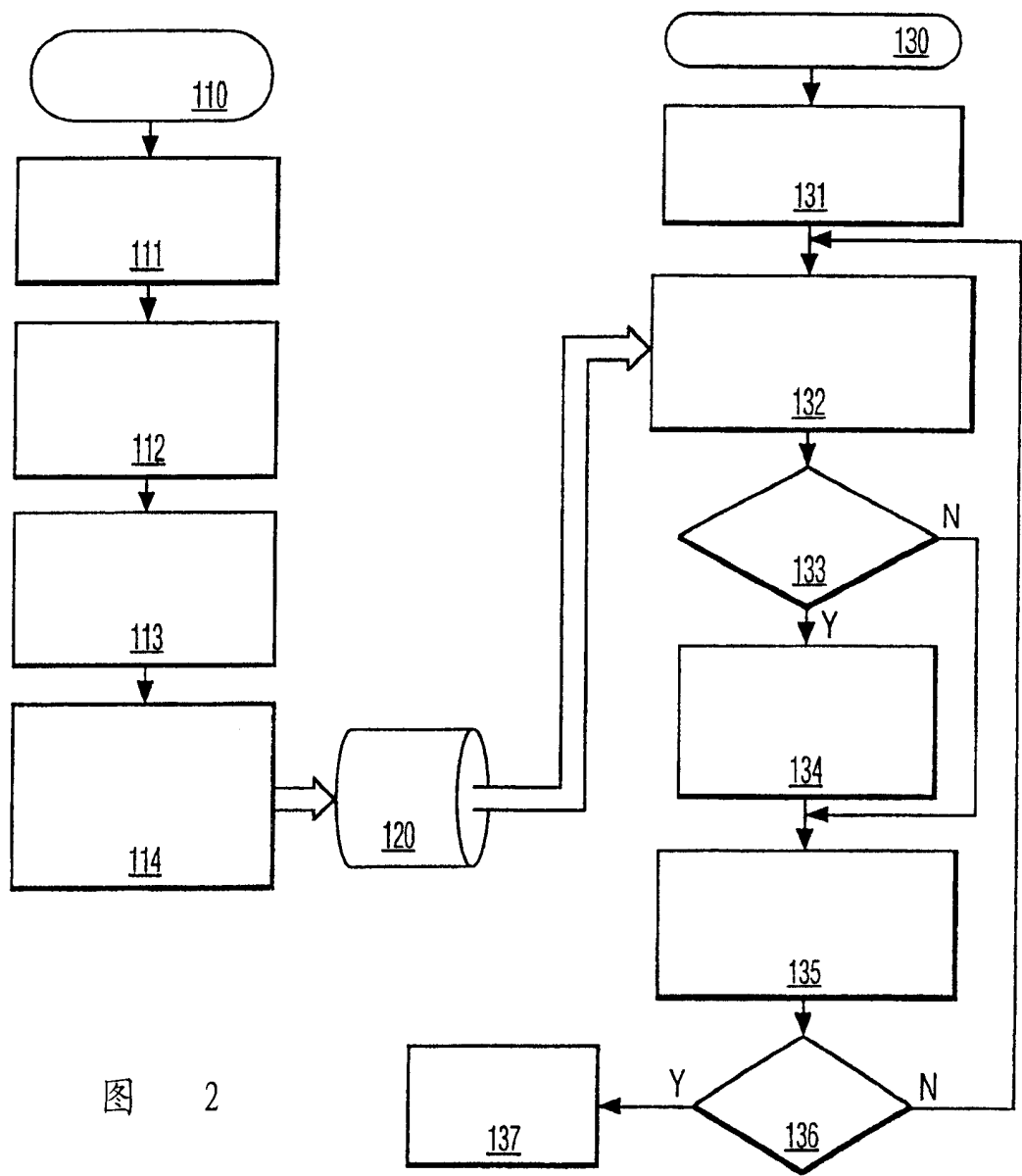


图 2

MM_INFO_NO	DV	MDV		
MM01	DV01	MDV01		
MM02		MDV02		
MM03	DV03	MDV03		
MM04	DV04	MDV04		
MM05				
MM06				
MM07		MDV07		
MM08	DV08	MDV08		

图 3

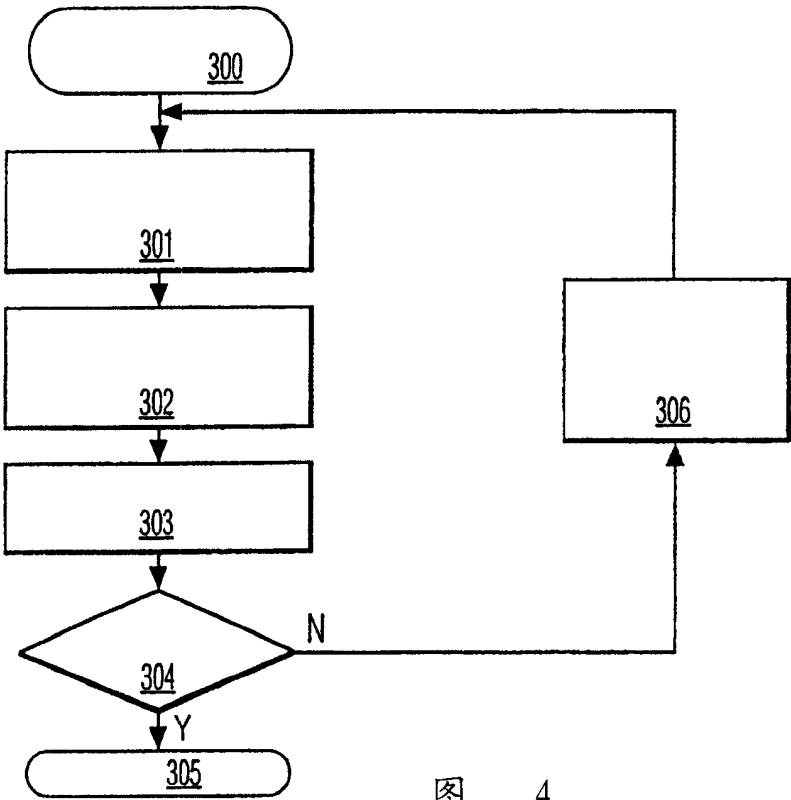


图 4