



(12) 发明专利申请

(10) 申请公布号 CN 102016839 A

(43) 申请公布日 2011. 04. 13

(21) 申请号 200980112906. X

(51) Int. Cl.

(22) 申请日 2009. 04. 10

G06F 17/30 (2006. 01)

(30) 优先权数据

12/102, 395 2008. 04. 14 US

(85) PCT申请进入国家阶段日

2010. 10. 12

(86) PCT申请的申请数据

PCT/US2009/002270 2009. 04. 10

(87) PCT申请的公布数据

W02009/131629 EN 2009. 10. 29

(71) 申请人 伊斯曼柯达公司

地址 美国纽约州

(72) 发明人 罗杰波 袁俊松

(74) 专利代理机构 北京集佳知识产权代理有限公司

公司 11227

代理人 李春晖 陈炜

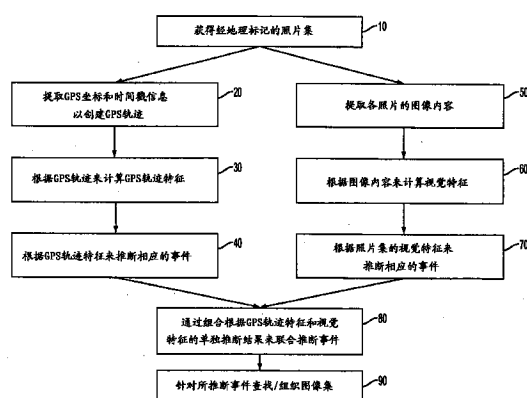
权利要求书 2 页 说明书 8 页 附图 7 页

(54) 发明名称

使用捕捉位置序列信息的图像分类

(57) 摘要

公开了对一组时间相关图像的分类, 其中, 根据该组时间相关图像来识别捕捉位置序列。使共同与捕捉位置序列相关联的捕捉位置序列信息与预定捕捉位置序列特征的多个集合中的每个集合进行比较。每个集合与预定分类相关联。至少基于比较步骤的结果来识别与该组时间相关图像相关联的所标识的分类; 以及把所识别出的分类存储在处理器可访问存储器系统中。



1. 一种至少部分地由数据处理系统实施的方法，所述方法用于便于一组时间相关图像的分类，所述方法包括以下步骤：

根据该组时间相关图像来识别捕捉位置序列；

使共同与所述捕捉位置序列相关联的捕捉位置序列信息与预定捕捉位置序列特征的多个集合中的每个集合进行比较，每个集合都与预定分类相关联；

至少基于所述比较步骤的结果来识别与该组时间相关图像相关联的所标识的分类；以及把所识别出的分类存储在处理器可访问存储器系统中。

2. 如权利要求 1 所述的方法，其中，至少基于与该组中的图像中的至少一些图像相关联的全球定位系统信息来识别所述捕捉位置序列。

3. 如权利要求 1 所述的方法，还包括以下步骤：把共同与该组时间相关图像相关联的图像内容信息与预定图像内容特征的多个集合中的每个集合相比较，预定图像内容特征的每个集合都与预定分类相关联，其中，所述识别步骤至少基于所述两个比较步骤的结果来识别所标识的分类。

4. 如权利要求 1 所述的方法，其中，根据所述预定分类中的一个分类中来识别所标识的分类。

5. 如权利要求 1 所述的方法，其中，所述捕捉位置序列信息包括所述捕捉位置序列的轨迹。

6. 如权利要求 5 所述的方法，其中，所述轨迹包括空间域和时间域。

7. 如权利要求 5 所述的方法，还包括以下步骤：确定所述轨迹的特征，所述特征涉及由所述轨迹表示的空间范围或由所述轨迹表示的移动速度，其中，所述比较步骤涉及使所述特征与在预定捕捉位置序列特征集合中的至少一个集合中表示的相应值或值的范围进行比较。

8. 如权利要求 5 所述的方法，还包括以下步骤：

确定所述轨迹的特征，所述特征涉及熵、距形心的距离、方差、偏度、峰度、空间范围、时间持续长短、长度或速度，

其中，所述比较步骤涉及使所述特征与在预定捕捉位置序列特征集合中的至少一个集合中表示的相应值或值的范围进行比较。

9. 如权利要求 1 所述的方法，其中，所述预定分类包括徒步旅行和城市观光。

10. 如权利要求 9 所述的方法，其中，所述预定分类还包括婚礼、球赛、后院活动、海滩活动、道路旅行、野餐、圣诞节活动以及生日会。

11. 一种存储指令的处理器可访问存储器系统，所述指令被配置为使得数据处理系统实施便于一组时间相关图像的分类的方法，其中，所述指令包括：

用于根据该组时间相关图像来识别捕捉位置序列的指令；

用于使共同与所述捕捉位置序列相关联的捕捉位置序列信息与预定捕捉位置序列特征的多个集合中的每个集合进行比较的指令，每个集合都与预定分类相关联；

用于至少基于所述比较步骤的结果来识别与该组时间相关图像相关联的所标识的分类的指令；以及

用于把所识别出的分类存储在处理器可访问存储器系统中的指令。

12. 如权利要求 11 所述的处理器可访问存储器系统，其中，所述捕捉位置序列信息包

括所述捕捉位置序列的轨迹。

13. 如权利要求 12 所述的处理器可访问存储器系统，其中，所述轨迹包括空间域和时间域。

14. 如权利要求 12 所述的处理器可访问存储器系统，进一步包括以下步骤：

确定所述轨迹的特征，所述特征涉及由所述轨迹表示的空间范围或由所述轨迹表示的移动速度，

其中，所述比较步骤涉及使所述特征与在预定捕捉位置序列特征的集合中的至少一个集合中表示的相应值或值的范围进行比较。

15. 如权利要求 12 所述的处理器可访问存储器系统，进一步包括以下步骤：

确定所述轨迹的特征，所述特征涉及熵、距形心的距离、方差、偏度、峰度、空间范围、时间持续长短、长度或速度，

其中，所述比较步骤涉及使所述特征与在预定捕捉位置序列特征的集合中的至少一个集合中表示的相应值或值的范围进行比较。

16. 一种系统，包括：

数据处理系统；以及

存储器系统，以通信方式连接到所述数据处理系统并且存储指令，所述指令被配置为使得所述数据处理系统实施便于一组时间相关图像的分类的方法，其中，所述指令包括：

用于根据该组时间相关图像来识别捕捉位置序列的指令；

用于使共同与所述捕捉位置序列相关联的捕捉位置序列信息与预定捕捉位置序列特性的多个集合中的每个集合进行比较的指令，每个集合都与预定分类相关联；

用于至少基于来自所述比较步骤的结果来识别与该组时间相关图像相关联的所标识的分类的指令；以及

用于把所识别出的分类存储在所述存储器系统中的指令。

17. 如权利要求 16 所述的存储器系统，其中，所述捕捉位置序列信息包括所述捕捉位置序列的轨迹。

18. 如权利要求 17 所述的存储器系统，其中，所述轨迹包括空间域和时间域。

19. 如权利要求 17 所述的存储器系统，进一步包括以下步骤：

确定所述轨迹的特征，所述特征涉及由所述轨迹表示的空间范围或由所述轨迹表示的移动速度，

其中，所述比较步骤涉及使所述特征与在预定捕捉位置序列特征的集合中的至少一个集合中表示的相应值或值的范围进行比较。

20. 如权利要求 17 所述的存储器系统，进一步包括以下步骤：

确定所述轨迹的特征，所述特征涉及熵、距形心的距离、方差、偏度、峰度、空间范围、时间持续长短、长度或速度，

其中，所述比较步骤涉及使所述特征与在预定捕捉位置序列特征的集合中的至少一个集合中表示的相应值或值的范围进行比较。

使用捕捉位置序列信息的图像分类

技术领域

[0001] 本发明涉及人工智能和图像理解，具体而言，涉及时间相关图像集的分类。

背景技术

[0002] 消费者照片集是普遍的。从这些集中在语义上发掘有意义的信息已成为机器学习和计算机视觉团体中积极研究的领域。存在关注于对象识别（诸如，检测某些类型的对象，如脸、车辆、草、水、天空等）问题的大量工作。大多数此工作依靠使用图像中可获得的低级视觉特征（诸如，颜色、纹理和线条）。近年来，越来越关注于从语义上提取更复杂的信息，诸如，场景检测和活动识别。例如，现有系统已尝试通过场景和对象的视觉分类来识别事件，参见例如 L.-J.Li 和 L.Fei-Fei, “What, where and who? classifying events by scene and object recognition”, Proc.IEEE Intl.Conf.on Computer Vision, 2007。该系统报告了在识别许多特殊体育事件（诸如，马球、划船以及博彩(boche)）方面的适度成功，这是由于可以从这些事件的图片中观测到的独特视觉特性。

[0003] 在所有以上提到的现有技术中，只使用基于图像的特征（例如，颜色和边缘直方图、或者“视觉特征袋”）以及基于各图像来执行传统图像聚类 and 分类（参见 S.Lazebnik、C.Schmid 以及 J.Ponce, “Beyond bags of features: spatial pyramid matching for recognizing natural scene categories”, Proc.IEEE Conf.on Computer Vision and Pattern Recognition, 2006）。

[0004] 然而，由于属于同一事件的图像之间的时间和空间相关性，所以图像常常并非是彼此独立的。更具体地，在个人图像集中，除了图像特征之外还存在丰富的上下文信息，这种上下文信息通常是对图像特征的补充以用于语义理解的目的。

[0005] 因此，需要考虑图像之间关系的改进图像分类技术。

发明内容

[0006] 通过根据本发明的各种实施例的用于便于一组时间相关图像分类的方法和系统在本领域中获得了技术方案并且解决了上述问题。在本发明的实施例中，根据该组时间相关图像来识别捕捉位置序列。使共同与捕捉位置序列相关联的捕捉位置序列信息与预定捕捉位置序列特性的多个集合中的每个集合进行比较。每个集合与预定分类相关联。至少基于来自比较步骤的结果来识别与该组时间相关图像相关联的所标识的分类；以及把所识别出的分类存储在处理器可访问存储器系统中。

[0007] 本发明的优点包括使用具有相关联捕捉位置信息的个人图像集来识别事件。

[0008] 除了上述实施例以外，通过参照附图以及通过对以下详细描述的学习应当清楚更多实施例。

附图说明

[0009] 图 1 和图 1a 是用于实施本发明实施例的流程图；

[0010] 图 2 是示出了在空间范围和移动速度方面不同事件的特性的表；

[0011] 图 3 和图 4 分别描绘了与城市观光事件和徒步旅行事件相关的空间和时间轨迹的示例；

[0012] 图 5 描绘了用于计算用来组合来自多个提示方式的识别结果的可信度的流程和示例混淆矩阵，以及

[0013] 根据本发明的实施例，图 6 示出了用于使用照片集和相关联的捕捉位置轨迹进行事件识别的系统。

具体实施方式

[0014] 共同推断可以是用于对图像进行分类的有用工具。例如，可能难以分辨特定图片是工作还是空闲，但是看着属于同一事件的其它图片，做出同样的预测会变得更加容易。在本发明的实施例中，在具有相关联捕捉位置信息的时间相关图像集的上下文中进行图像分类。本文中所使用的用语“图像分类”是指把一组时间相关图像分类成行为的预定集合中的一个行为的过程。进一步地，本文中所使用的用语“事件识别”和“行为识别”是图像分类的示例。更进一步地，词语“图像”、“图片”以及“照片”在本文中可互换使用。

[0015] 本发明包括本文中描述的实施例的组合。对“特定实施例”等的引用是指在本发明至少一个实施例中出现的特征。对“实施例”或“特定实施例”等的单独引用并非必定是指同一实施例或同样的实施例；然而，这些实施例并非是相互排他的，除非如此表明或者对于本领域技术人员而言是显而易见的。在引用“方法”等时使用单数和/或复数并非是限制性的。

[0016] 另外，应当注意，除非上下文另行明确注明或需要，在本公开中以非排他性的方式使用词语“或”。

[0017] 当把 GPS（全球定位系统）传感器安装在数字摄像装置中时，可以针对每个捕捉到的图像收集图像捕捉的时间、经度和纬度。另外，可以使用本领域中已知的技术针对每个捕捉到的图像捕捉以下信息，低等级图像特征：例如，颜色和边缘直方图；语义对象和场景识别：来自图像分类器（室内/室外检测器、海滩检测器等）和对象检测器的输出（例如，面部）；以及其它上下文信息：大致在相同时间和相同地点拍摄的图像（例如，同一事件或同一图像文件夹中的图像）。

[0018] 根据本发明的实施例，图 6 示出了用于事件识别的系统 100。系统 100 包括数据处理系统 110、外围系统 120、用户接口系统 130 以及处理器可访问存储器系统 140。处理器可访问存储器系统 140、外围系统 120 以及用户接口系统 130 以通信方式连接到数据处理系统 110。

[0019] 数据处理系统 110 包括实施本发明各种实施例的过程（包括图 1 和图 1a 的示例过程）的一个或更多个数据处理设备。用语“数据处理设备”或“数据处理器”旨在包括任何数据处理设备，诸如，中央处理单元（“CPU”）、桌上型计算机、膝上型计算机、大型计算机、个人数字助理、黑莓（Blackberry™）、数字摄像装置、手机、或者用于处理数据、管理数据、或者操控数据的任何其它设备或其组件，无论是否用电、磁、光、生物组件实施。

[0020] 处理器可访问存储器系统 140 包括被配置成存储信息（包括执行本发明各种实施例的过程（包括本文中描述的图 1 和图 1a 的示例过程）所需要的信息）的一个或多个处理器可访问存储器。处理器可访问存储器系统 140 可以是分布式处理器可访问存储器系统，包括经由多个计算机和 / 或设备以通信方式连接到数据处理系统 110 的多个处理器可访问存储器。另一方面，处理器可访问存储器系统 140 无需是分布式处理器可访问存储器系统，因此，可以包括位于单个数据处理器或设备内的一个或多个处理器可访问存储器。

[0021] 用语“处理器可访问存储器”旨在包括任何处理器可访问数据存储设备，无论是易失性的还是非易失性的、电子的、磁的、光的或其它，包括但不限于：寄存器、软盘、硬盘、压缩光盘、DVD、闪存、ROM 以及 RAM。

[0022] 用语“以通信方式连接”旨在包括可以传送数据的程序、数据处理器、或者设备之间的任何类型的连接，无论是有线的还是无线的。进一步地，用语“以通信方式连接”旨在包括单个数据处理器内的程序或设备之间的连接、位于不同数据处理器中的程序或设备之间的连接、以及根本不位于数据处理器中的设备之间的连接。在此方面，虽然处理器可访问存储器系统 140 被示出为与数据处理系统 110 分开，但本领域技术人员应当理解，可以把处理器可访问存储器系统 140 完全地或部分地存储在数据处理系统 110 内。在此方面进一步地，虽然外围系统 120 和用户接口系统 130 被示出为与数据处理系统 110 分开，但本领域技术人员应当理解，可以把这些系统中的一个系统或二者完全地或部分地存储在数据处理系统 110 内。

[0023] 外围系统 120 可以包括被配置成向数据处理系统 110 提供数字图像的一个或多个设备。例如，外围系统 120 可以包括数字视频摄像装置、手机、常规数字摄像装置或者其它数据处理器。数据处理系统 110 在从外围系统 120 中的设备接收到数字内容记录后可以把这些数字内容记录存储在处理器可访问存储器系统 140 中。

[0024] 用户接口系统 130 可以包括鼠标、键盘、另一计算机或者用于向数据处理系统 110 输入数据的任何设备或设备的组合。在此方面，虽然外围系统 120 被示出为与用户接口系统 130 分开，但外围系统 120 可以被包括作为用户接口系统 130 的一部分。

[0025] 用户接口系统 130 也可以包括显示设备、处理器可访问存储器或者用于由数据处理系统 110 向其输出数据的任何设备或设备的组合。在此方面，如果用户接口系统 130 包括处理器可访问存储器，则这种存储器可以是处理器可访问存储器系统 140 的一部分，尽管在图 6 中分开示出了用户接口系统 130 和处理器可访问存储器系统 140。

[0026] 本发明便于图像分类，诸如，人拍摄图片和视频的典型事件的自动识别。目标是从个人照片集（一些可能包含视频）中识别典型事件，其中，每个事件对应于在某个环境中发生、并且被在事件期间拍摄的图像集记录的特定人类行为：

[0027] $E_i = \{I_j : j = 1, 2, \dots, n\}$

[0028] 其中， I_j 表示图像（或视频）。在本发明的优选实施例中，定义了以合理区分的视觉特性的 10 种类型频繁发生的事件或类别的列表： $C = \{\text{后院活动, 海滩活动, 球赛, 生日会, 圣诞节活动, 城市观光, 徒步旅行, 道路旅行, 滑雪以及婚礼}\}$ 。可以定义和包括其它类型的事件，本领域技术人员将会理解，本发明不限于这些事件。列表 C 包括室外和室内事件。通常，由于涉及语义的更高等级——视觉内容可以从一个示例到另

一示例显著变化以及使得分类器表现不佳，所以事件识别比场景识别更具挑战性和复杂（参见 L.-J.Li 和 L.Fei-Fei, “What, where and who? classifying events by scene and object recognition”, Proc.IEEE Intl.Conf.on Computer Vision, 2007.）。

[0029] 可以把本发明实施例解决的分类型问题叙述如下：

[0030] 输入：

[0031] 视觉事件的数据库被输入作为用于设计分类器的训练数据 D , $D = \{E_i, l(E_i)\}$, 其中, $E_i = \{T_i, j, I_i, j: j = 1, 2, \dots, |E_i|\}$ 表示包含构建视觉事件的图像集合的单个图像文件夹（组）并且 $|E_i|$ 表示 E_i 的大小； $T_i, j = (x_i, j, y_i, j, t_i, j)$ 是照片的 2 维位置 (x, y) 和时间 t 的 GPS 记录, I_i, j 是相应的图像。共同地, 所有记录 T_i, j 形成轨迹 T_i 。对于每个图像文件夹, 可以用 C 中视觉事件类中的一个 $l(E_i)$ 来唯一地标注该图像文件夹, 其中, $C = \{\text{后院活动, 海滩活动, 球赛, 生日会, 圣诞节活动, 城市观光, 徒步旅行, 道路旅行, 滑雪以及婚礼}\}$ 。所有图像 I_i, j 共享与它们所属的组 E_i 相同的标签。训练数据集 D 应该包含 C 中每个类的足够示例。

[0032] 输出：

[0033] 标签 $l(E_i)$ 给出了新图像文件夹 $E_i = \{T_i, j, I_i, j: j = 1, 2, \dots, |E_i|\}$ 和训练后的分类器, 例如, $l(E_i) = \text{徒步旅行}$ 。

[0034] 在任何模式分类问题中, 需要把高维特征向量分类成预定义类集合中的一个类。由于特征之间的冗余以及训练样本的有限数量, 常常难以在初始高维空间中找到单个好分类器（最有可能是非线性分类器）。相反, 更容易在低维特征空间中找到许多弱分类器, 其中, 发现的子空间中每个构造出的分类器作为弱分类规则。在本发明的实施例中, 有效的方法是通过多类提升 (boosting) 来组合这些在低维空间中发现的弱分类器以创建最终分类器。提升是本领域中公知的并且可以在不偏离本发明范围的情况下使用用于创建最终分类器的其它类似流程。

[0035] 在图 1 中, 根据本发明的实施例, 示出了用于对一组时间相关图像进行分类的方法。首先, 在步骤 10 获得一组具有捕捉位置信息的时间相关图像。该组或这些图像（在本文中有时称作集）被根据本领域中的已知技术分成了或者分成时间和位置相关照片组。另外, 可以使用装备有 GPS 接收传感器的数字摄像装置对该图像集进行地理标记 (geotag)。或者, 可以使用能够（例如通过蓝牙链接）与独立 GPS 接收器通信的数字摄像装置来获取经地理标记的照片。其它备选方案用于那些要在捕捉之后被手动标记有坐标、或者按批与 GPS 记录设备同步的照片。在步骤 20 中, 对于与特定事件相对应的经地理标记的照片集, 首先提取 GPS 坐标和时间戳信息以形成空间和时间上的稀疏轨迹。轨迹被视为捕捉位置序列。注意, 并非每个照片都必需与位置信息相关联, 只要存在足够多的具有地点信息的照片以形成整组时间相关图像的轨迹即可。

[0036] 接下来, 在步骤 30 中计算轨迹特征以表征轨迹的形状或结构。在步骤 40 中使用这些轨迹特征来推断针对该集的相应事件或分类。通过把共同与捕捉位置序列相关联的信息与期望的捕捉位置序列特性的多个集合中的每个集合相比较来执行推断, 每个集合与事件的分类相关联。

[0037] 除了使用捕捉位置序列信息（诸如, 步骤 20-40 中的位置序列轨迹）之外, 为了对该集进行分类, 可选地可以使用该集中的图像内容信息, 如步骤 50-70 所反映的那

样。特别地，可以在步骤 50 中从每个单个照片中提取图像内容，即，像素信息，随之是步骤 60，其中根据图像内容来计算视觉特征。可以根据共同由同一事件中所有的各照片表示的视觉特征来在步骤 70 中推断事件或分类。再次地，通过把共同与该组时间相关图像相关联的图像内容信息与期望的图像内容特性的多个集合中的每个集合相比较来执行推断，每个集合与事件的分类相关联。因为存在两个并行的推断路径，所以在步骤 80 中可以通过组合根据轨迹特征和视觉特征单独获得的推断结果来联合推断事件。

[0038] 最后，可以在步骤 90 中使用所推断的事件信息来在大数据库中查找特定事件的照片，或者注释照片集用于其它方面。

[0039] 图 1a 是在步骤 85 中根据组合的轨迹特征和视觉特征（即，组合信息）联合推断事件的备选方法，没有根据轨迹特征和视觉特征单独推断的中间步骤。

[0040] 若干示例情形可以示出本发明如何比只使用视觉信息的系统表现得更好。假定一组人进行了道路旅行并且经过了主要城市。还假定他们在城市中和道路上拍摄了图片。仅使用视觉信息，视觉分类器很可能将会把此事件中的许多图片分类成“城市观光”但是很可能将会误表征整个事件。然而，使用与图片相关联的捕捉位置序列轨迹，整组图片有更大的机会被正确地分类成道路旅行（尽管经过城市）。在信息融合之后，最终的判定是此事件是道路旅行（注意，本发明学习了道路旅行事件可以包含城市的图片）。

[0041] 另一示例将会是一组人在树林中进行了徒步旅行。由于后院图片也可以包含树木，所以视觉分类器会错误地判定这是后院事件。然而，再次使用与图片相关联的捕捉位置序列轨迹，本发明的实施例有更大的机会正确地确定整组事件是徒步旅行，而不是后院事件。

[0042] 与只可应用视觉特征的传统图像和视觉事件分类问题不同，本发明的实施例采用两种类型的特征：(1)GPS 特征和 (2) 视觉特征。特别地，在与同一事件相关的图像集上定义 GPS 特征，因此，GPS 特征是整個图像集而非只是个别图像的特性。在本发明的上下文中，照片集也被称作文件夹。

[0043] 图 1 中提到的稀疏轨迹对于区分各种事件可以确实是有用的。如图 2 中所总结的，简单地根据移动的速度和空间范围，可以把事件分成不同的簇，如此前所定义的那样这些簇中的每个簇与标签相关联。本发明的实施例可以假定已根据本领域中已知的技术把照片预先分组成时间和位置相关照片的组。事件（诸如城市观光、徒步旅行以及道路旅行）覆盖较大的空间范围，而后院、海滩、球赛、野餐以及婚礼事件倾向于是空间上受局限的。而在道路旅行中，人以相当高的速度行进，对于城市观光、徒步旅行、后院、海滩、球赛、野餐以及婚礼而言移动更加缓慢。可以设计比范围和速度更复杂的特征以表征轨迹的形状或结构。

[0044] 基于所记录的 GPS 信息，每个图像文件夹由空间时间轨迹表示： $T_i = \{(x_i, j, y_{ij}, t_i, j)|, j = 1, \dots, |T_i|\}$ ，其中， $|T_i|$ 等于文件夹的大小，即，图像的数量。空间时间域中点的序列提供了对摄像者在图像捕捉过程中经过的基本移动的指示。与在空间时间上的这些点处拍摄的相应图像独立的此序列可独自为事件识别的任务提供信息。图 3 和图 4 分别包含若干城市观光事件和徒步旅行事件的 GPS 轨迹。可以在用于指定分类器的训练数据 D 中使用这些轨迹，在上述分类问题的“输入”部分中提及了。值得注

意，这两种类型事件的轨迹的形状特性彼此明显不同，而同时，同样类型事件的轨迹看起来相似。虽然两个事件倾向于在空间和时间上覆盖相对而言较大的范围，但城市观光轨迹因为潜在的城市街区和街道结构而在轨迹的形状上呈现出更高程度的规则性。相比而言，徒步旅行轨迹即使它们也可以是 Z 字形的也看起来更随机和随意。换言之，同样事件类型的不同移动示例之间存在结构相似性，而对应于不同事件类型的移动之间存在结构差别。

[0045] 给定这种空间时间轨迹，在本发明的优选实施例中，在步骤 50 中提取具有要通过训练确定并且在 [0.0, 1.0] 的范围中的变化重要性量级而不失一般性的以下 13 种类型的特征，作为轨迹形状或结构的表示：

[0046] 熵 (2 个特征)：时间和空间域

[0047] 距形心的距离 (1 个特征)

[0048] 方差 (2 个特征)：时间和空间域

[0049] 偏度 (2 个特征)：时间和空间域

[0050] 峰度 (2 个特征)：时间和空间域

[0051] 空间范围 (1 个特征) 和时间持续长短 (1 个特征)

[0052] 轨迹长度 (1 个特征) 和平均速度 (1 个特征)

[0053] 例如，下面是以上列出的空间范围和空间方差的定义：(1) 轨迹 T_i 的空间范围 $SR(T_i)$ ，描述视觉事件的活动空间范围： $SR(T) = (\max\{x_j | x_j \in T\} - \min\{x_j | x_j \in T\}) \times (\max\{y_j | y_j \in T\} - \min\{y_j | y_j \in T\})$ ；(1) 以及 (2) 轨迹的空间方差 $SV(T_i)$ ，描述活动的空间方差：

$$[0054] \quad SV(T) = \frac{1}{2}[VAR(\{x_j | x_j \in T\}) + VAR(\{y_j | y_j \in T\})]; \quad (2)$$

[0055] 其它类型的特征是根据图像在步骤 60 中计算的低等级视觉特征。已证明了视觉词汇是建立视觉识别系统的有效方式，例如，用于场景识别（见 S.Lazebnik、C.Schmid 以及 J.Ponce，“Beyond bags of features：spatial pyramid matching for recognizing natural scene categories”，Proc.IEEE Conf.on Computer Vision and Pattern Recognition, 2006）。图像被通过固定网格分区以及表示成图像块的无序集合。对这些图像块计算合适的描述以及把这些描述聚类到区块 (bin) 中以形成“视觉词汇”。在本发明的实施例中，同样的方法学可以被采用和扩展以考虑用于表征每个图像网格的颜色和纹理特征。

[0056] 在本发明的实施例中，为了提取颜色特征，把图像网格进一步分区成 2×2 个相等大小的子网格。随后对于每个子网格，提取平均 R、G 和 B 值以形成表征 4 个子网格颜色信息的 $4 \times 3 = 12$ 的特征向量。为了提取纹理特征，计算每个子网格中具有 8 个朝向区块的直方图的 2×2 的阵列。从而应用 $4 \times 8 = 32$ 维的 SIFT 描述符来表征每个图像网格内的结构，在原理上与 Lazebnik 类似。在本发明的实施例中，如果图像大于 200k 个像素，则首先把它调整为 200k 个像素。随后在重叠 8×8 的采样间隔的情况下设置 16×16 的图像网格大小。通常，一个图像生成 117 个这种网格。

[0057] 在从图像网格中提取所有原始图像特征之后，在本发明的实施例中，通过如下方式建立单独的颜色和纹理词汇：通过 k 均值 (k-means) 聚类在训练数据集中对所有图像网格进行聚类。把两种词汇均设置为大小 500。通过累积事件中的所有网格 (图像集)，对事件获得两个归一化的直方图，即，hc 和 ht，分别对应于颜色和纹理词汇的词语

分布。联接 hc 和 ht, 结果是归一化的词语直方图: $\sum_{i=1}^{1000} h_i(E) = 1$ 。直方图中的每个区块表明相应词语的出现频率。因此特征向量 f 具有响应:

[0058]

$$f_i(E) = \begin{cases} 1 & \text{如果 } h_i(E) > \theta_i \\ 0 & \text{其它,} \end{cases} \quad (3)$$

[0059] 其中, $0 < \theta_i < 1$ 是量化阈值。考虑简化指数函数, 即, fi, 当事件中第 i 个词语 Ω 足够频繁时给出肯定响应, 否则空响应。存在两种类型的视觉词语 Ω_c 和 Ω_t , 其中, Ω_c 是颜色原词, 而 Ω_t 是纹理原词。把完整词语表示成 $\Omega = \Omega_c \cup \Omega_t$, 可以计算视觉特征以表示颜色和纹理信息。

[0060] 除了低等级视觉特征以外, 还可以考虑从对象和场景检测得出的语义视觉特征。对于每个图像, 可以定义场景类别的集合:

[0061] $S = \{\text{室外, 人造, 海滩, 日落, 山, } \dots\}$ 以及对象类别的集合:

[0062] $O = \{\text{草地, 蓝天, 沙, 雪, } \dots\}$

[0063] 可以对这些场景和对象建立二进制 SVM 分类器的集合是众所周知的, 例如参见 Lazebnik 中。替代做出确定性的判定, 本发明的优选实施例维持范围在 $[0, 1]$ 之间的概率分值以表明图像是否包含特定对象或场景。注意, 允许为每个图像分配多个场景或对象标签。因此, 对于每个图像 I, 结果是 k 维实值特征向量, k 维实值特征向量的每个值都在 $[0, 1]$ 内, 其中, k 等于场景和对象类别之和并且每个值对应于来自各个 SVM 分类器的检测分值。

[0064] 在计算视觉特征之后, 可以使用被称为具有指数损失的逐级相加建模 (SAMME) 的多类 AdaBoost 方法来使用训练数据集 D 中的图像建立视觉分类器、以及使用训练集 D 中的捕捉位置序列轨迹建立捕捉位置序列轨迹分类器。可以在 (J.Zhu、S.Rosset、H.Zou 以及 T.Hastie, “Multi-class Adaboost”, Technical Report, 2005, 斯坦福大学) 中找到 SAMME 的详细内容。SAMME 通过建立单个 K 类分类器而非 K 个二进制分类器直接处理 K 类问题。SAMME 的方案与贝叶斯分类规则一致, 因此在使误分类误差最小化方面是最优的。

[0065] 普通的融合方案 (例如, 提升) 是使用许多弱分类器的线性组合以创建强分类器。最终的判定是:

[0066]

$$\hat{H}(D_q) = \sum_{i=1}^L w_i h_i(D_q), \quad (4)$$

[0067] 其中, w_i 是需要优化的权重, $h_i(D_q) \in \{0, 1\}$ 表示来自弱分类器的响应 $h_i(D_q)$ 。

[0068] 在针对轨迹特征、低等级视觉特征以及语义视觉特征中的每个都建立了单独分类器之后, 本发明就通过信息融合来组合这些并行分类的结果。对于简单的融合方案, 可以例如针对 8 个事件类进行计算:

[0069] GPS 输出: $hg(X) = [4.2 \ 0.0 \ 0.5 \ 4.9 \ 0.0 \ 0.0 \ 0.0 \ 0.7]$

[0070] 视觉输出: $hv(X) = [1.2 \ 0.7 \ 4.1 \ 1.7 \ 3.1 \ 0.0 \ 2.0 \ 0.0]$

[0071] - 输出的和： $h(X) = h_g(x) + h_v(X)$ (5)

[0072] $h(X) = [5.4 \ 0.7 \ 5.6 \ 6.6 \ 3.1 \ 0.0 \ 2.0 \ 0.7]$

[0073] 在此情形中，预测的标签 = $\arg \max h(X) = 4$ 。

[0074] 在本发明的优选实施例中，通过基于置信度的融合来考虑每个类的可靠性，类似于使用贝叶斯网络的概率融合，即，

[0075] $h(X) = W_g \times h_g(X) + W_v \times h_v(X)$ (6)

[0076] 可以基于相应的分类模式（轨迹、视觉或者语义）的混淆矩阵计算权重或置信度。混淆矩阵是本领域中众所周知的并且表示如何把每个类（每行中）的样本分类成可能类（每列）中的每个类。因此，好的分类器应该具有大多数非零值沿着矩阵的对角线聚集的混淆矩阵。对某个模式的 8 类问题使用图 5 中的示例混淆矩阵，针对此模式的第一类的权重被计算成

[0077] $W_g(1) = 5 / (5+2+4+0+1+2+0+1) = 1/3$ (7)

[0078] 其中，分母是第一列之和。简而言之，从分类为该模式中的类 1 的 15 个样本中，5 的确在类 1 中，所以置信度是 1/3。

[0079] 特别参照某些优选实施例详细描述了本发明，但是应当理解，可以在本发明的精神和范围内进行改变和修改。

[0080] 组成部分列表

[0081] 10 步骤：获取经地理标记的照片集

[0082] 20 步骤：从经地理标记的照片提取 GPS 坐标及其时间戳信息以创建轨迹

[0083] 30 步骤：根据轨迹来计算轨迹特征

[0084] 40 步骤：根据轨迹特征来推断相应的事件

[0085] 50 步骤：提取各照片的图像内容

[0086] 60 步骤：根据图像内容来计算视觉特征

[0087] 70 步骤：根据照片集的视觉特征来推断相应的事件

[0088] 80 步骤：通过组合根据轨迹特征和视觉特征的单独推断结果来联合推断事件

[0089] 85 根据轨迹特征和视觉特征来联合推断事件

[0090] 90 步骤：查找 / 组织用于所推断事件的图像集

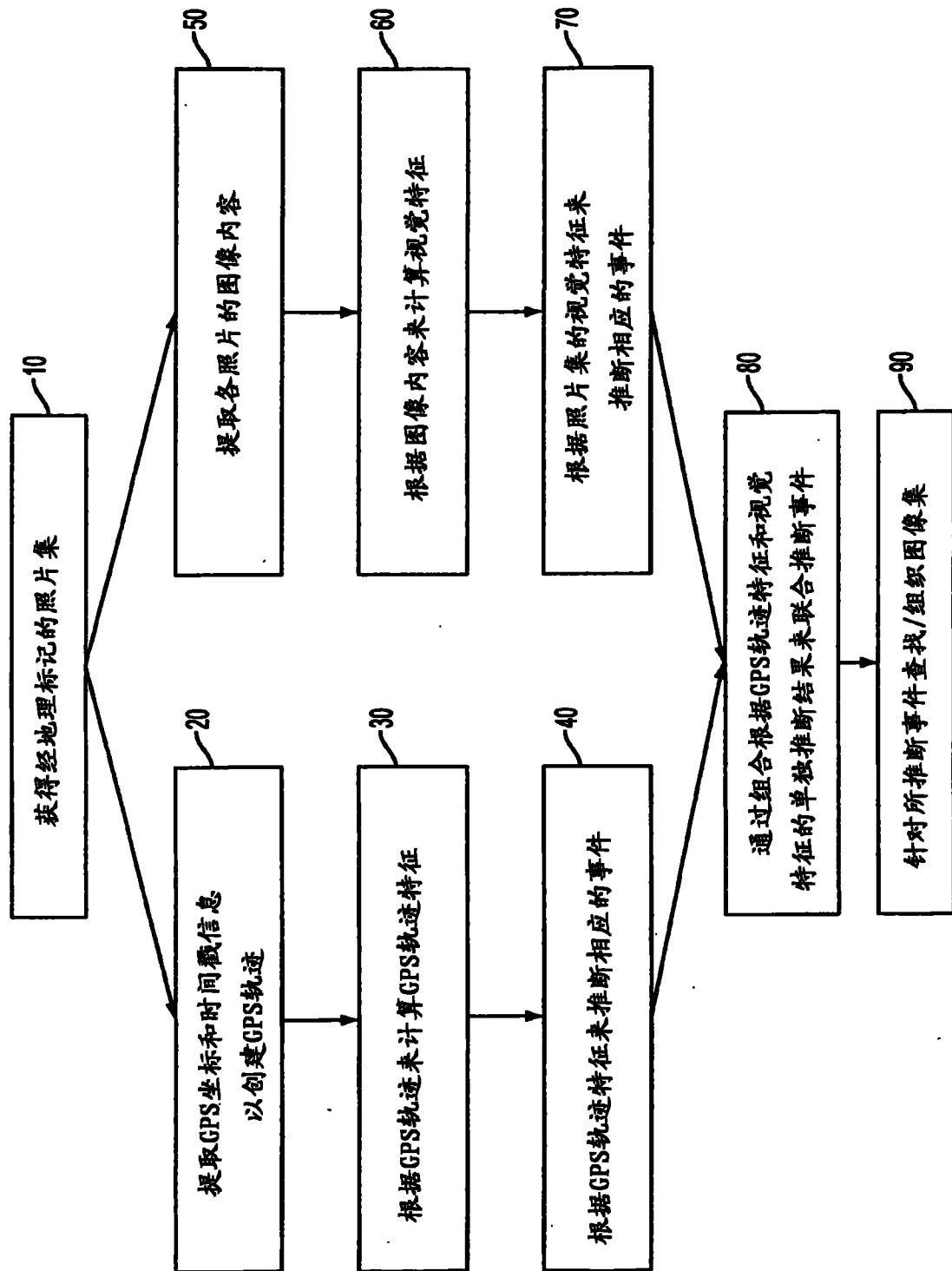


图 1

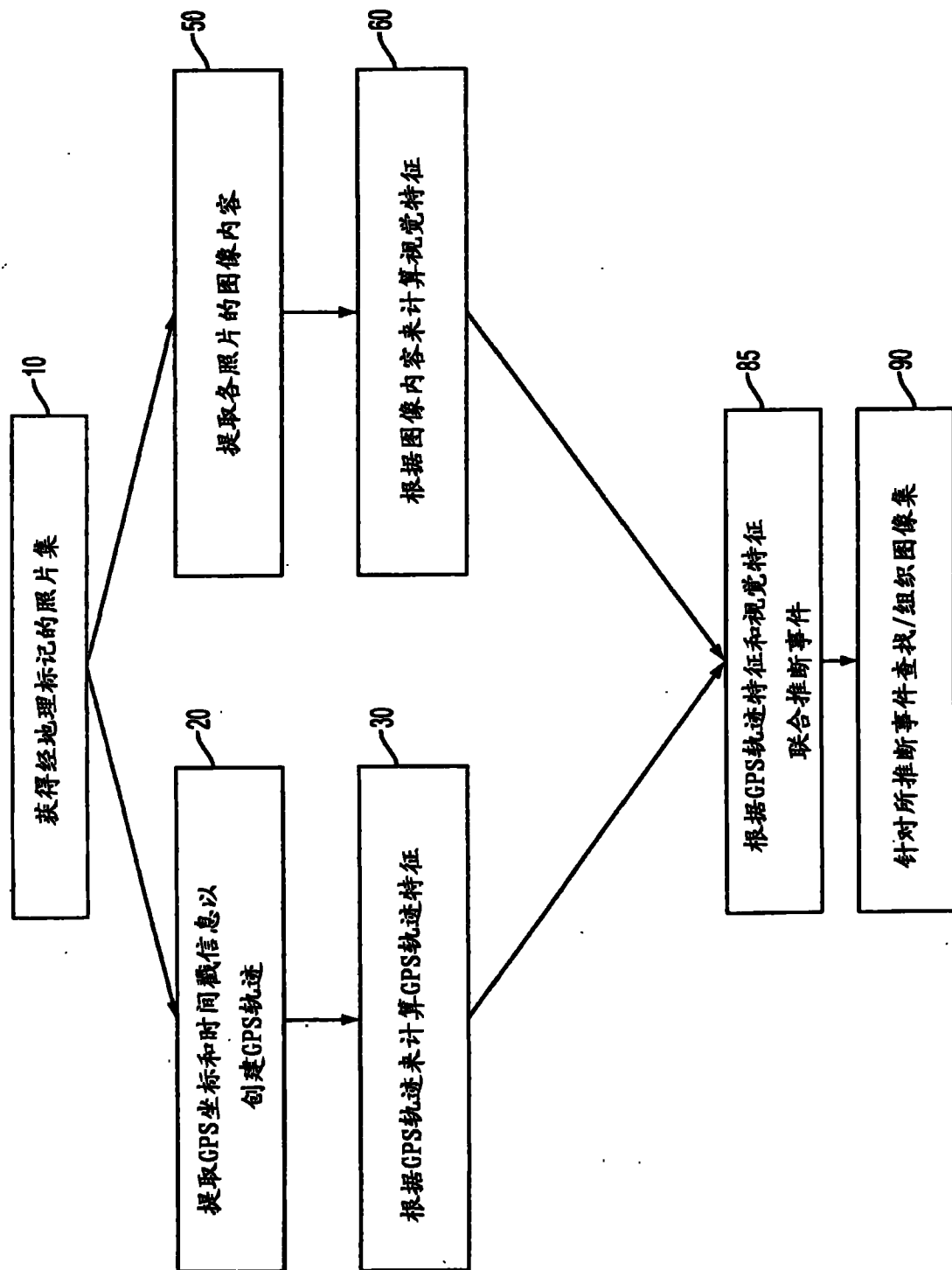


图 1A

大空间范围的事件	小空间范围的事件
城市观光、徒步旅行、道路旅行	后院、海滩、球赛、野餐、婚礼

高移动速度的事件	低至中移动速度的事件
道路旅行	城市观光、徒步旅行、后院、海滩、球赛、野餐、婚礼

图 2

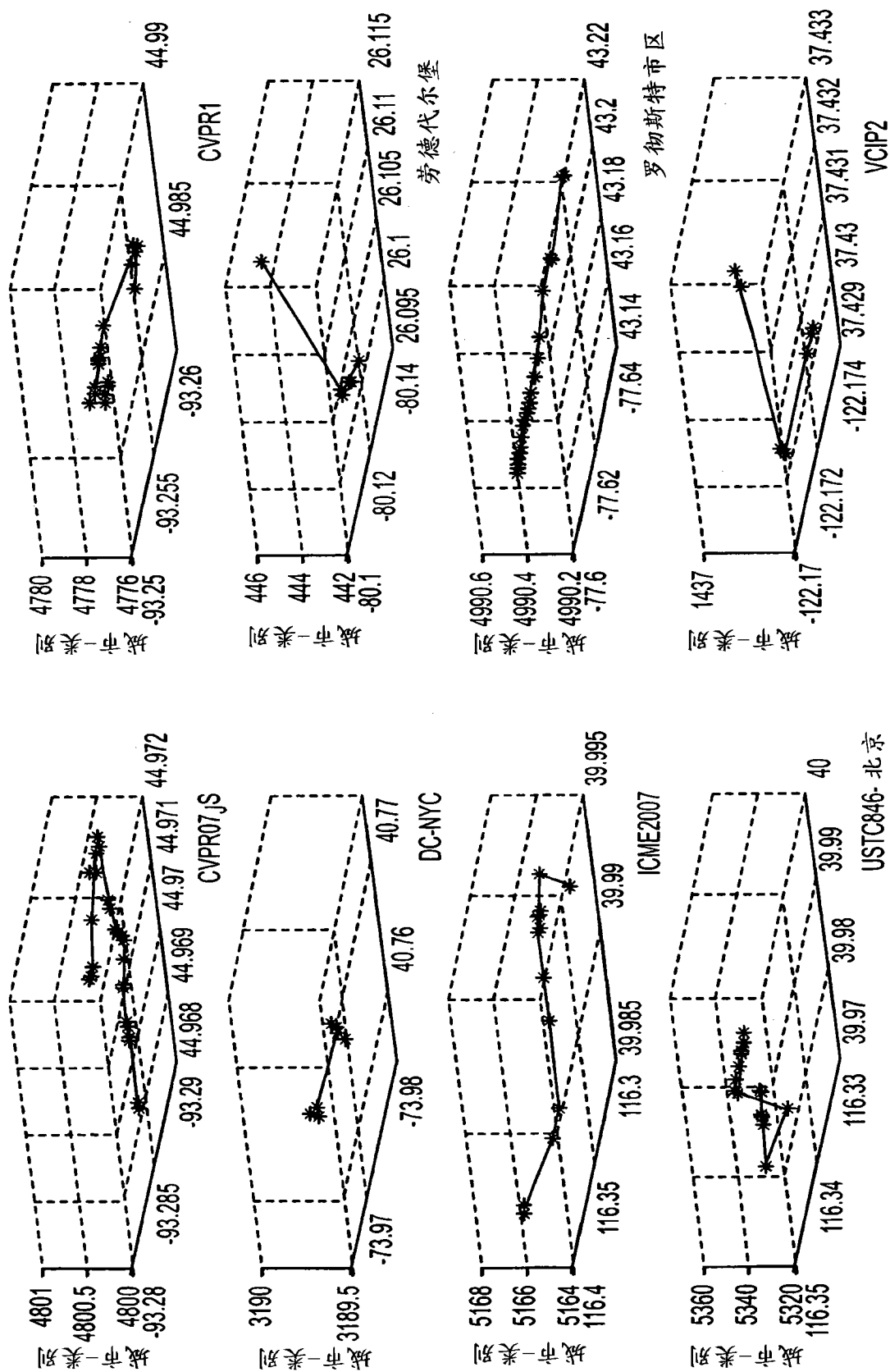


图 3

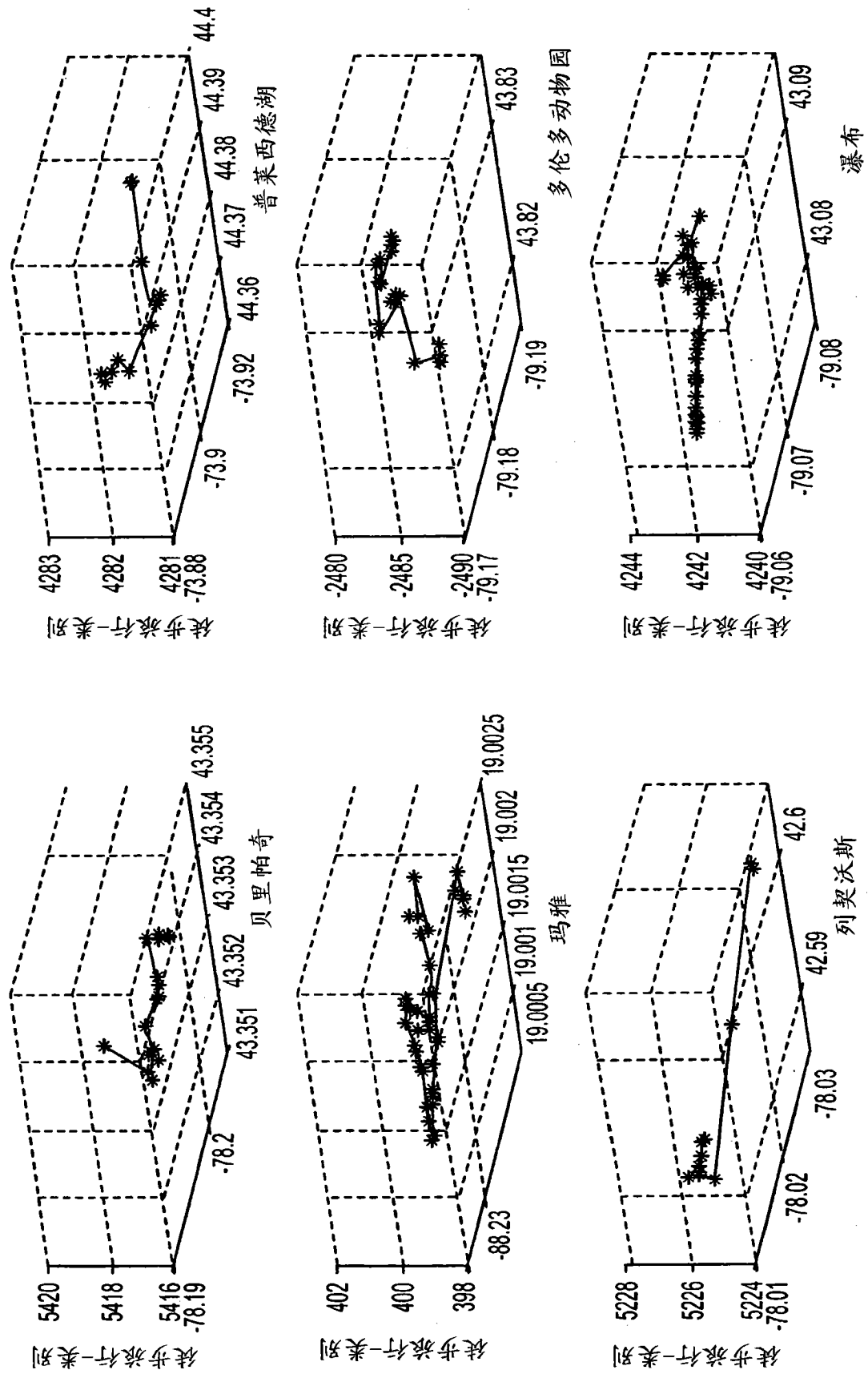


图 4

GPS特征混淆矩阵

5	0	2	0	0	1	0	0
2	2	0	1	0	1	0	1
4	1	0	0	0	0	0	0
0	0	0	7	1	0	0	0
1	0	0	1	5	0	0	0
2	2	1	0	0	0	0	0
0	0	0	0	0	0	8	0
1	2	0	0	0	0	0	0

$$W_g(1) = 5/(5+2+4+0+1+2+0+1) = 1/3$$

图 5

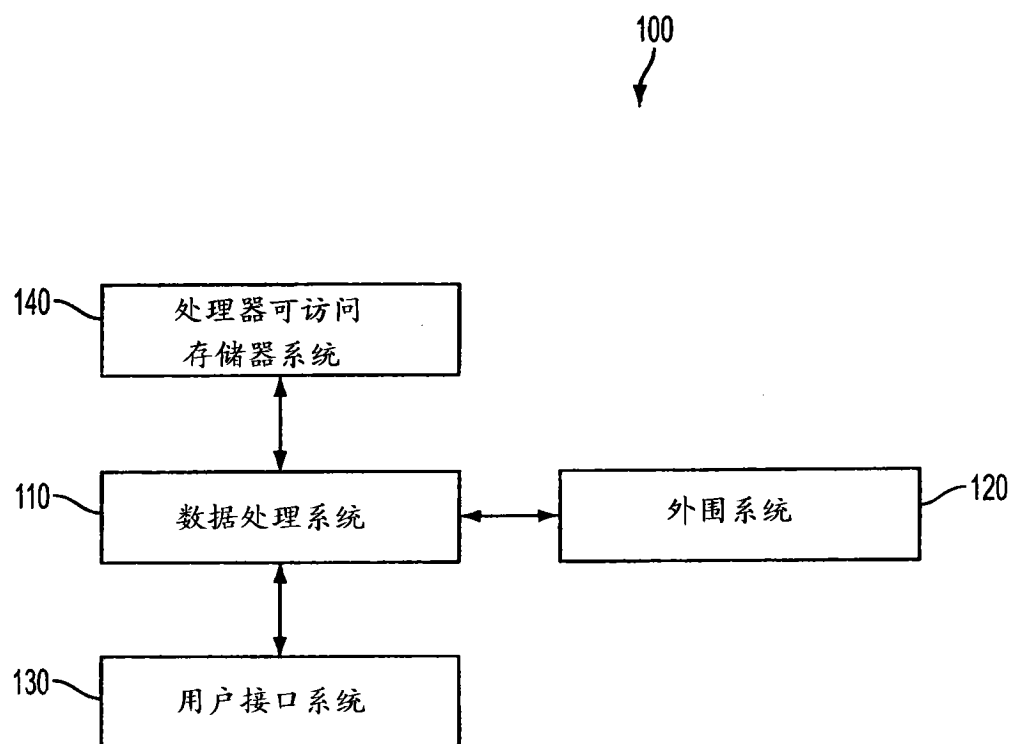


图 6