



(12) 发明专利申请

(10) 申请公布号 CN 103324677 A

(43) 申请公布日 2013.09.25

(21) 申请号 201310199867.X

(22) 申请日 2013.05.24

(71) 申请人 西安交通大学

地址 710049 陕西省西安市咸宁西路 28 号

(72) 发明人 李婧 钱学明

(74) 专利代理机构 西安通大专利代理有限责任  
公司 61200

代理人 徐文权

(51) Int. Cl.

G06F 17/30 (2006.01)

G06K 9/54 (2006.01)

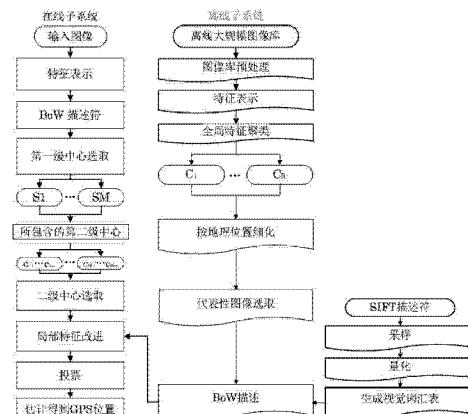
权利要求书2页 说明书10页 附图2页

(54) 发明名称

一种可分级的快速图像 GPS 位置估计方法

(57) 摘要

本发明公开了一种可分级的快速图像 GPS 位  
置估计方法，其对于离线的数据库，首先按照颜色  
纹理全局特征对图像进行全局特征聚类，然后，对  
聚类得到的图像类按照其拍摄的地理位置分到第  
二级的小中心中，最后对每个聚类中心进行代表  
性图像选取。在线系统中，首先根据颜色纹理特征  
选择第一级的候选类，然后在第一级候选类包含  
的第二级小中心中继续根据颜色纹理特征计算距  
离并选择第二级候选类，最后通过局部特征确认  
的方法最终对输入图像进行 GPS 位置估计。本发  
明的方法不仅能够有效提高图像的检索速度，而  
且能够改进检索精度。



1. 一种可分级的快速图像 GPS 位置估计方法,其特征在于,包括以下步骤:

(1) 离线图像库的处理

1) 图像库预处理

计算图像的平均亮度和 HWVP 纹理能量,按照计算的图像平均亮度,分别去除 1% 最高和 1% 最低的图像;按照计算的 HWVP 纹理能量,分别去除 1% 最高和 1% 最低的图像;

2) 图像特征提取

提取图像中的全局特征:45-D 颜色矩特征,170-D HWVP 特征还有局部的 SIFT 特征;所述 45-D 颜色矩特征是指将图像平均分成四份,再重叠提取图像中心区域,然后对五个区域使用九维颜色矩进行描述;

3) 全局特征聚类

按照全局特征,采用 K-means 聚类方法将图像库中的所有图像聚类成 32 个第一级聚类;

4) 根据地理位置细化

将上一步得到的 32 个第一级聚类中的每一类按照地理位置进行细分,得到新的第二级聚类;所述每个第二级聚类中图片都拍摄于相同的地理位置;

5) 代表性图片选取

对每个第二级聚类进行代表性图片选取,得到每个第二级聚类的代表性图像组;

6) 建立快速的文件索引结构

对代表性图像组使用视觉词汇包模型进行描述,并且构建视觉词汇与代表性图像组的倒排索引结构;

(2) 在线图像地理位置估计

1) 第一级中心选取

首先将输入图像的全局特征与第一级聚类的各个中心进行比对,选择距离最近的 M 个中心作为候选中心;其中比对是通过计算距离得到的,距离计算方法如下,

$$D_i = ||LC_i - L_{input}||, (i=1, \dots, R),$$

其中  $L_{input}$  表示输入图像的 45-D 颜色矩特征和 170-D HWVP 特征构成的 215 维全局特征,  $D_i$  表示输入图像与第  $i$  个类  $C_i$  全局特征中心  $LC_i$  的距离;  $||X||$  用来表示 X 的范数;  $R=32$ ;

按照计算的距离选择 M 个距离最近的第一级聚类作为候选中心  $S=\{S_1, \dots, S_M\}$ , 其中  $M \leq R$ ;

2) 第二级中心选择

在选择了 M 个第一级聚类候选中心  $S=\{S_1, \dots, S_M\}$  之后,进一步进行更精细的第二级聚类候选中心的选择;用  $s=\{r_1, \dots, r_N\}$  表示所选  $S=\{S_1, \dots, S_M\}$  中的所有的第二级聚类中心,其中  $r_i \in \{c_{j,k}\}$ ,  $j=1, \dots, R$ ;  $k=1, \dots, N_j$ ; ;然后在所有的 N 个第二级聚类中心中选择前 V% 作为选择出的第二级候选中心, V 为 0-100 的实数;选择方法与第一级聚类候选中心相同,计算方法如下:

$$d_i = ||Lr_i - L_{input}||, i \in \{1, \dots, N\}$$

其中 N 是所有第一级聚类候选中心下面所包含的第二级聚类中心的数目,  $Lr_i$  代表第  $i$  个第二级聚类候选中心的全局特征描述;

通过该步骤得到的所有的第二级聚类候选中心,记做  $SC = \{g_1, \dots, g_F\}$ , 其中  $g_f \in \{r_1, \dots, r_N\}$ , 其中  $f \in \{1, \dots, F\}$ ,  $F = V \times N / 100$ ;

### 3) 使用局部特征改进

使用 BoW 直方图或者倒排索引结构对选取的第二级聚类候选中心进行确认;再根据确认得到的结果使用 KNN 的方法对输入图像的地理位置进行推荐。

2. 根据权利要求 1 所述的可分级的快速图像 GPS 位置估计方法,其特征在于,当使用 BoW 直方图对选取的第二级聚类候选中心进行确认时,按照以下进行:

在离线系统中,用 BoW 模型描述之后,生成 BoW 直方图;在此,对输入图像也生成相对应维数的 BoW 直方图,并表示为  $h(k)$ ,  $k = 1, \dots, Q$ ;使用四种直方图距离度量方式进行距离度量,包括余弦距离 COS、城市距离 MAD、欧式距离 MSD 和直方图相交 HIST;其距离计算方法分别如下:

$$COS(i, j) = \frac{\sum_{k=1}^Q NH_{i,j}(k) \times h(k)}{\sqrt{\left[ \sum_{k=1}^Q (NH_{i,j}(k))^2 \right] * \left[ \sum_{k=1}^Q (h(k))^2 \right]}} \quad (6)$$

$$MAD(i, j) = \sum_{k=1}^Q |NH_{i,j}(k) - h(k)| \quad (7)$$

$$MSD(i, j) = \sum_{k=1}^Q (NH_{i,j}(k) - h(k))^2 \quad (8)$$

$$HIST(i, j) = \sum_{k=1}^Q \min(NH_{i,j}(k), h(k)) \quad (9)$$

其中, NH 是代表性图像组的 BoW 直方图描述。

3. 根据权利要求 1 所述的可分级的快速图像 GPS 位置估计方法,其特征在于,当使用倒排索引结构对选取的第二级聚类候选中心进行确认时,按照以下进行:

使用 TF-IDF 计算权值的方法,对每个图像组计算其相应的得分,得分计算方法如下,

$$Score(L) = \sum_{x=1}^Q \frac{\omega_x * Freq_L(x)}{Number_L * Frequency_x} \quad (10)$$

其中  $Freq_L(x)$  是第  $x$  号词汇的频数,  $Number_L$  是图像中的所有视觉词汇的数目,  $Frequency_x$  是词汇在整个图像库中出现的次数,  $\omega_x$  表示的是对于输入图像词汇的权重;权重计算方法如下,

$$\omega_x = \frac{Freq_{input}(x)}{Number_{input}} \quad (11)$$

其中  $Freq_{input}(x)$  表示 BoW#x 的频次 and  $Number_{input}$  输入图像所包含的的视觉词汇的总数。

## 一种可分级的快速图像 GPS 位置估计方法

### 技术领域

[0001] 本发明属于多媒体图像处理技术领域，涉及一种图像位置估计方法，尤其是一种可分级的快速图像 GPS 位置估计方法，特别涉及图像内所包含景物的位置识别。

### 背景技术

[0002] 随着生活水平的提高，人们开始热衷旅游，并在旅游过程中拍摄大量图片上传到图片分享网站与社交网络。知名的图片分享网站如 Flickr，上传的图片总数达 50 亿张。社交网络中的图片上传量更为惊人，光 Facebook 一家就达到了 600 亿。在我国，社交网站人人网，开心网成为了主要的上传与分享的途径。与此同时，人们也越来越方便的收集各种景点的各种图片。也许人们会碰到自己非常喜欢的图像，但是却不知道这些图像是在什么地方拍的。因此，对于如此大规模图像多媒体数据，如何有效地利用这些巨大规模的多媒体数据帮助人们自动的识别图像中包含的景物的拍摄地点是很有必要的。

[0003] 目前的图像识别一般都要借助图像所包含的一些标签或者评论信息，根据标签和评论信息对图像进行分类与识别。基于文本的图像检索也已经日渐成熟，像谷歌、百度等，给定输入关键词的前提下，能都搜索出一系列的带有关键词标签的图像。但是由于图像附带的标签与描述存在噪声，另外，这些附带信息本身是人们添加上去的，难免会有一些主观性。所以，利用图像内容进行图像检索以及图像的识别是非常有必要的，另一方面，网络上的大规模图像多媒体资源，也为自动的通过图像内容进行图像的识别提供了有利的条件。

[0004] 随着科技的进步，人们智能手机和部分数码相机拍照时带有时间和 GPS 记录功能，在上传不自带 GPS 信息的图片时，Flickr 提供用户可以手动拖动上传图像到其相应位置的能。因此，是可以通过网络多媒体手段获取大量的带有地理位置标签的图像的。这就对自动的图像地理位置估计提供了条件。

### 发明内容

[0005] 本发明的目的在于提供一种可分级的快速图像 GPS 位置估计方法，该方法利用两个子系统：离线子系统及在线子系统。其中离线子系统主要对大规模带有地理位置标记图像库的处理。而对于在线子系统，是对查询图像进行特征描述，使用全局特征对所有的离线图库中的图像类进行筛选，之后根据全局特征选取候选图像组并且使用局部特征进行特征量化得到视觉词汇，最后结合每个图像类的代表性图库的倒排索引表，最后根据距离，对结果进行排序，并且使用 K-NN 方法对图像地理位置进行估计。在这个过程中，使用了倒排索引结构达到了快速图像检索的目的，通过快速的图像检索能够加快图像 GPS 位置估计的完成。该方法能够提高图像的检索速度，改进检索精度。

[0006] 本发明的目的是通过以下技术方案来解决的：

[0007] 这种可分级的快速图像 GPS 位置估计方法，包括以下步骤：

[0008] (1) 离线图像库的处理

[0009] 1) 图像库预处理

- [0010] 计算图像的平均亮度和 HWVP 纹理能量,按照计算的图像平均亮度,分别去除 1% 最高和 1% 最低的图像;按照计算的 HWVP 纹理能量,分别去除 1% 最高和 1% 最低的图像;
- [0011] 2) 图像特征提取
- [0012] 提取图像中的全局特征:45-D 颜色矩特征,170-D HWVP 特征还有局部的 SIFT 特征;所述 45-D 颜色矩特征是指将图像平均分成四份,再重叠提取图像中心区域,然后对五个区域使用九维颜色矩进行描述;
- [0013] 3) 全局特征聚类
- [0014] 按照全局特征,采用 K-means 聚类方法将图像库中的所有图像聚类成 32 个第一级聚类;
- [0015] 4) 根据地理位置细化
- [0016] 将上一步得到的 32 个第一级聚类中的每一类按照地理位置进行细分,得到新的第二级聚类;所述每个第二级聚类中图片都拍摄于相同的地理位置;
- [0017] 5) 代表性图片选取
- [0018] 对每个第二级聚类进行代表性图片选取,得到每个第二级聚类的代表性图像组;
- [0019] 6) 建立快速的文件索引结构
- [0020] 对代表性图像组使用视觉词汇包模型进行描述,并且构建视觉词汇与代表性图像组的倒排索引结构;
- [0021] (2) 在线图像地理位置估计
- [0022] 1) 第一级中心选取
- [0023] 首先将输入图像的全局特征与第一级聚类的各个中心进行比对,选择距离最近的 M 个中心作为候选中心;其中比对是通过计算距离得到的,距离计算方法如下,
- [0024]  $D_i = ||LC_i - L_{input}||$ , ( $i=1, \dots, R$ ),
- [0025] 其中  $L_{input}$  表示输入图像的 45-D 颜色矩特征和 170-D HWVP 特征构成的 215 维全局特征,  $D_i$  表示输入图像与第  $i$  个类  $C_i$  全局特征中心  $LC_i$  的距离; $||X||$  用来表示 X 的范数; $R=32$ ;
- [0026] 按照计算的距离选择 M 个距离最近的第一级聚类作为候选中心  $S=\{S_1, \dots, S_M\}$ , 其中  $M \leq R$ ;
- [0027] 2) 第二级中心选择
- [0028] 在选择了 M 个第一级聚类候选中心  $S=\{S_1, \dots, S_M\}$  之后,进一步进行更精细的第二级聚类候选中心的选择;用  $s=\{r_1, \dots, r_N\}$  表示所选  $S=\{S_1, \dots, S_M\}$  中的所有的第二级聚类中心,其中  $r_i \in \{c_{j,k}\}$ ,  $j=1, \dots, R$ ;  $k=1, \dots, N_j$ ; ;然后在所有的 N 个第二级聚类中心中选择前 V% 作为选择出的第二级候选中心, V 为 0-100 的实数;选择方法与第一级聚类候选中心相同,计算方法如下:
- [0029]  $d_i = ||Lr_i - L_{input}||$ ,  $i \in \{1, \dots, N\}$
- [0030] 其中 N 是所有第一级聚类候选中心下面所包含的第二级聚类中心的数目,  $Lr_i$  代表第  $i$  个第二级聚类候选中心的全局特征描述;
- [0031] 通过该步骤得到的所有第二级聚类候选中心,记做  $SC = \{g_1, \dots, g_F\}$ , 其中  $g_f \in \{r_1, \dots, r_N\}$ , 其中  $f \in \{1, \dots, F\}$ ,  $F = V \times N / 100$ ;
- [0032] 3) 使用局部特征改进

[0033] 使用 BoW 直方图或者倒排索引结构对选取的第二级聚类候选中心进行确认；再根据确认得到的结果使用 KNN 的方法对输入图像的地理位置进行推荐。

[0034] 进一步，以上当使用 BoW 直方图对选取的第二级聚类候选中心进行确认时，按照以下进行：

[0035] 在离线系统中，用 BoW 模型描述之后，生成 BoW 直方图；在此，对输入图像也生成相对应维数的 BoW 直方图，并表示为  $h(k)$ ,  $k = 1, \dots, Q$ ；使用四种直方图距离度量方式进行距离度量，包括余弦距离 COS、城市距离 MAD、欧式距离 MSD 和直方图相交 HIST；其距离计算方法分别如下：

$$[0036] COS(i, j) = \frac{\sum_{k=1}^Q NH_{i,j}(k) \times h(k)}{\sqrt{\left[ \sum_{k=1}^Q (NH_{i,j}(k))^2 \right] * \left[ \sum_{k=1}^Q (h(k))^2 \right]}} \quad (6)$$

$$[0037] MAD(i, j) = \sum_{k=1}^Q |NH_{i,j}(k) - h(k)| \quad (7)$$

$$[0038] MSD(i, j) = \sum_{k=1}^Q (NH_{i,j}(k) - h(k))^2 \quad (8)$$

$$[0039] HIST(i, j) = \sum_{k=1}^Q \min(NH_{i,j}(k), h(k)) \quad (9)$$

[0040] 其中，NH 是代表性图像组的 BoW 直方图描述。

[0041] 进一步的，以上当使用倒排索引结构对选取的第二级聚类候选中心进行确认时，按照以下进行：

[0042] 使用 TF-IDF 计算权值的方法，对每个图像组计算其相应的得分，得分计算方法如下，

$$[0043] Score(L) = \sum_{x=1}^Q \frac{\omega_x * Freq_L(x)}{Number_L * Freq_{input_x}} \quad (10)$$

[0044] 其中  $Freq_L(x)$  是第  $x$  号词汇的频数， $Number_L$  是图像中的所有视觉词汇的数目， $Freq_{input_x}$  是词汇在整个图像库中出现的次数， $\omega_x$  表示的是对于输入图像词汇的权重；权重计算方法如下，

$$[0045] \omega_x = \frac{Freq_{input}(x)}{Number_{input}} \quad (11)$$

[0046] 其中  $Freq_{input}(x)$  表示 BoW#x 的频次 and  $Number_{input}$  输入图像所包含的的视觉词汇的总数。

[0047] 与现有技术相比，本发明优点在于：

[0048] 1) 本发明的可分级的快速图像 GPS 位置估计方法使用视觉特征和地理位置信息对大规模的图像库构建了可分层的结构，能够有效提高图像的检索速度；

[0049] 2) 本发明提出了分层聚类和局部特征改进的 GPS 地理位置估计方法，通过结合局部特征，能够有效改进检索精度；

[0050] 3) 本发明在局部特征的使用过程中,引用了倒排索引结构来加快检索速度和提高检索准确度。

## 附图说明

[0051] 图 1 为本发明方法的总体步骤示意框图;

[0052] 图 2 为本发明方法中采用倒排索引机构的时候的倒排索引表的结构图。

## 具体实施方式

[0053] 下面结合附图对本发明做进一步详细描述:

[0054] 参见图 1 :本发明的可分级的快速图像 GPS 位置估计方法包括以下步骤:

[0055] (1) 离线图像库的处理

[0056] 离线子系统的目的是对大规模带地理标签的图像库进行离线处理,使其能够更好的应用于在线的图像地理位置估计,以达到快速有效的进行 GPS 估计的目的。离线子系统主要包括以下六个方面:1) 图像库的预处理,2) 对图像库内图像的全局和局部特征的描述,3) 使用全局特征对图像库种图像进行 K-means 聚类;4) 根据地理位置对第一级的聚类得到的类进行细分得到第二级的图像组,5) 对第四步得到的图像组进行代表性图像的选择,6) 对代表性图像组建立倒排索引表。下面分别对这六个部分进行介绍。

[0057] 1) 图像库预处理

[0058] 图像库预处理的目的在于去除一些噪声图像。因为在图像库构建的过程中,是根据关键词在图像库中检索然后进行下载的。这样就导致下载到的图像质量很不均一,可能会有很多图像亮度特别高 / 低,或者包含特别多的噪声,然后这些图像对于完成 GPS 估计是没有作用,甚至是反作用的,于是本发明中就先进行了初步的预处理,旨在去除那些质量比较差的图像。按照计算的图像平均亮度,分别去除 Alpha% 最高和最低的图像,通过后面实验发现 Alpha 取 1 的时候效果最好。对于包含噪声多纹理复杂的图像的去除,本发明中使用的是 HWVP 特征,计算了纹理能量,然后进行噪声的去除。通过预处理,保证了剩余参与离线图像库使用的图像是一些质量比较好的图。因此,在本发明的最佳方案中,该步骤按照以下进行:

[0059] 计算图像的平均亮度和 HWVP 纹理能量,按照计算的图像平均亮度,分别去除 1% 最高和 1% 最低的图像(即将图像平均亮度排序,去掉两头高部分的和低部分的各 1%) ;按照计算的 HWVP 纹理能量,分别去除 1% 最高和 1% 最低的图像;

[0060] 2) 图像特征提取

[0061] 提取图像中的全局特征:45-D 颜色矩特征,170-D HWVP 特征还有局部的 SIFT 特征;所述 45-D 颜色矩特征是指将图像平均分成四份,再重叠提取图像中心区域,然后对五个区域使用九维颜色矩进行描述;

[0062] 3) 全局特征聚类

[0063] 考虑到大规模图像库图像检索中对检索速度的要求,本发明中提出使用聚类中心代替图像库中的单幅图像的思想。因为相比图像数,中心数能够减低很多,这样所需要比较的时间也就缩短了。通过全局特征聚类,整个图像库就能够划分成一些相对规模较小的图像类,而每个图像类都有相似的全局特性。之所以聚类,是基于减低复杂度的考虑。另外也

希望通过全局特征聚类,能够将拍摄于同一个地方的图像,能够按不同的场景进行划分,如春夏秋冬,白天和黑夜,地标与非地标建筑,古典建筑与现代建筑。

[0064] 按照全局特征,采用 K-means 聚类方法将图像库中的所有图像聚类成 32 个第一级聚类;

[0065] 对于全局特征中使用的特征包括颜色特征和纹理特征。颜色特征使用的是 45 维的颜色矩,而纹理特征则是使用的 170 维的 HWVP 特征。本发明中将两个特征融合成一个 215 维的低级特征,对图像库中所有的图像分别进行描述。通过聚类,得到 R 个类  $C_1, \dots, C_R$ ,并且每个类都有一个中心  $LC_i$ 。

[0066] 4) 根据地理位置细化

[0067] 本步骤中,将上一步得到的 32 个第一级聚类中的每一类按照地理位置进行细分,得到新的第二级聚类;所述每个第二级聚类中图片都拍摄于相同的地理位置。具体解释如下:

[0068] 得到了第一级的聚类  $r_i \in \{c_{j,k}\}$ ,  $j = 1, \dots, R$ ;  $k = 1, \dots, n$ ;之后,根据图像的拍摄地点对这 R 个聚类进行更精细的划分。假设第 i 个中心  $C_i$  下面包含拍摄于  $N_i$  ( $i = 1, \dots, R$ ) 个景点的图像,那么进行按照地理位置的进一步细分,得到了第二级的图像聚类  $c_{i,j}$  ( $i = 1, \dots, R$ ;  $j = 1, \dots, N_i$ )。对于每一个第二级的聚类,本文进行了其全局特征中心的计算,并用  $LC_{i,j}$  表示。其计算公式如下所示,

$$[0069] LC_{i,j} = \frac{1}{n_{i,j}} \sum_{k=1}^{n_{i,j}} L_{i,j,k}, \quad i = 1, \dots, R; j = 1, \dots, N_i \quad (1)$$

[0070] 其中  $n_{i,j}$  是  $c_{i,j}$  中所包含图像的数目,  $L_{i,j,k}$  表示  $c_{i,j}$  中第 k 张图像的 215 维的全局特征向量。第二级图像数目与第一级图像数目之间的关系如下所示,

$$[0071] z_i = \sum_{j=1}^{N_i} n_{i,j}; \quad i = 1, \dots, R; j = 1, \dots, N_i \quad (2)$$

[0072] 因此,图像库中图像总数就是所有第一级中心内所有图像数目的加和,表示如下:

$$[0073] Z = \sum_{i=1}^R z_i; \quad i = 1, \dots, R; j = 1, \dots, N_i \quad (3)$$

[0074] 也就是说,通过聚类和按照 GPS 位置,对图像库进行了划分,每个图像都被分到了相应的第一级和第二级的中心中,也就有了相应的第一级类别和第二级类别。

[0075] 5) 代表性图片选取

[0076] 对每个第二级聚类进行代表性图片选取,得到每个第二级聚类的代表性图像组。

[0077] 得到了第二级聚类  $c_{i,j}$  之后,所能够保证的是所有类内的图像都是具有相似的全局特征以及拍摄于相似的地方。由于第二级中心的精细划分是根据图像所附带的标签进行的,而网络上的很多标签都是人工标记包含大量噪声的,因此,  $c_{i,j}$  中的图像并不能保证确实是拍摄于所标记的地点中。另外一个方面,由于拍摄于相同景点的图像可能会有不同的拍摄角度,也就导致会有较大的差异。所以仅仅通过全局中心来表示整个  $c_{i,j}$  就导致由于取了均值带来的内部差异的隐藏。考虑到局部特征描述符能够有效的提取图像的局部信息,能够挖掘出图像中所包含的景点,本发明采用 SIFT 特征匹配的方法对每一个第二级的聚类进行代表性图像的选取。选取的具体算法如下,该算法是参考了 Kennedy 等人文章

中的代表性图像选择的办法改进进行选择的。算法 1 是本发明提出的代表性图像选择的流程：

---

**Algorithm 1: Selecting Representative Images For the GPS Location Refined Cluster**


---

 $c_{ij}$ **Input:**All the images in  $c_{ij}$  denoted set  $D$ 

[0078]

**Initial:**Pair-wise **match between every pair of images in  $D$ :****Determine** the number of images matching each image in  $D$  by counting the number of matching SIFT features.**Remove** images without matched image from  $D$ ; $A \leftarrow$  the image with most matched images in  $D$ ;**Update:**  $l \leftarrow 1$ ,  $\Omega_l \leftarrow A$ ,  $D \leftarrow D - A$ **while**  $D$  is not null $A \leftarrow$  the image with most matched images in  $D$ ; $P_A \leftarrow$  Number of SIFT features in  $A$ ;**for**  $k=1:l$ Count the number of matched SIFT point  $n_k$  between image  $A$  and image  $\Omega_k$ ;**end**

[0079]

 $P_* \leftarrow \max\{n_1, \dots, n_l\}$ ,  $* \leftarrow \arg \max_k \{n_k\}$ **if**  $P_* > P_A/2$ **then** image  $A$  can be viewed as near duplicate with image  $\Omega_*$ .**update:**  $D \leftarrow D - A$ **otherwise** image  $A$  is assigned as a representative image for the centroid,**update:**  $l \leftarrow l+1$ ;  $\Omega_l \leftarrow A$ ;  $D \leftarrow D - A$ **end****Output:** representative images  $\{\Omega_1, \dots, \Omega_l\}$  for the GPS location refined centroid  $c_{ij}$ 

[0080] 6) 建立快速的文件索引结构

[0081] 对代表性图像组使用视觉词汇包模型进行描述，并且构建视觉词汇与代表性图像组的倒排索引结构。

[0082] 在图像多媒体检索领域，表示局部视觉特征内容的描述符或者局部特征区域的向量就可以看做是一个个的视觉词汇。比如出现在不同图像中的相同的局部特征点 / 区域应该具有相同的属性，而且描述相同局部特征点临域或者局部特征区域的描述符应该也是一样的，这样可以说，两个图像的相应两个关键帧都拥有该视觉词汇。那么类比文本倒排索引的定义，图像库中所有图像所提取的局部特征描述符就可以看做是文档中的一个个词汇，而不同的图像就可以看做是不同的文档，那么基于视觉词汇的倒排索引其实就是要为每一个视觉词汇记录它出现的视频名称，出现的帧号甚至在该帧中的坐标位置。所以对于视觉词汇的倒排索引的建立来说，最重要的就是将数量无限的视觉词汇量化到数量有限的可以表示内容的视觉词汇上，我们也形象地称之为量化。量化之后就可以将图像使用 BoW 模型进行标示。

[0083] 参见图 2，对比传统的倒排表不同，本发明中的倒排表是词汇于每个代表性图像组之间的对应关系。倒排表中记录的是每个视觉词汇，以及其所出现的代表性图像组的标号。

[0084] (2) 在线图像地理位置估计

[0085] 本发明的在线图像地理位置估计是于在线子系统中完成的。在线子系统具体的模块可以由图 1 中的在线部分构成。与离线子系统相似,它的各个模块也可以按照实现功能的不同分为特征提取,第一级中心选取,第二级中心选取以及局部特征确认三大部分:

[0086] 1) 第一级中心选取

[0087] 首先将输入图像的全局特征与第一级聚类的各个中心进行比对,选择距离最近的 M 个中心作为候选中心;其中比对是通过计算距离得到的,距离计算方法如下,

[0088]  $D_i = ||LC_i - L_{input}||$ , ( $i=1, \dots, R$ ),

[0089] 其中  $L_{input}$  表示输入图像的 45-D 颜色矩特征和 170-D HWVP 特征构成的 215 维全局特征,  $D_i$  表示输入图像与第  $i$  个类  $C_i$  全局特征中心  $LC_i$  的距离;  $||X||$  用来表示 X 的范数;  $R=32$ ;

[0090] 这样就可以计算得到输入图像于所有第一级中心的聚类,然后本发明中对所有的距离按照升序的方式进行排序。由于图像之间的距离可以描述视觉相似性,距离越近,则表明视觉上面相似性越高。然后按照计算的距离选择 M 个距离最近的第一级聚类作为候选中心  $S=\{S_1, \dots, S_M\}$ ,其中  $M \leq R$ ;此处的选择方法有一下两个优点,相比计算与图像库中所有图像的距离,仅仅计算中心的距离相比而言就大大降低了计算量。而选择 M 个中心而非一个中心,是考虑到此步骤仅仅是粗略的选择一些候选图像集,而并非最终的精细选择,因此多选几个聚类有利于寻找到真正的代表输入图像的图像类。通过第一步的选择,得到了 M 个第一级候选中心,记为  $s = \{S_1, \dots, S_M\}$ 。

[0091] 2) 第二级中心选取

[0092] 在选择了 M 个第一级聚类候选中心  $S=\{S_1, \dots, S_M\}$  之后,进一步进行更精细的第二级聚类候选中心的选择;用  $s=\{r_1, \dots, r_N\}$  表示所选  $S=\{S_1, \dots, S_M\}$  中的所有的第二级聚类中心,其中  $r_i \in \{c_{j,k}\}$ ,  $j=1, \dots, R$ ;  $k=1, \dots, N_j$ ; ;然后在所有的 N 个第二级聚类中心中选择前 V% 作为选择出的第二级候选中心, V 为 0-100 的实数;选择方法与第一级聚类候选中心相同,计算方法如下:

[0093]  $d_i = ||Lr_i - L_{input}||$ ,  $i \in \{1, \dots, N\}$

[0094] 其中 N 是所有第一级聚类候选中心下面所包含的第二级聚类中心的数目,  $Lr_i$  代表第  $i$  个第二级聚类候选中心的全局特征描述;

[0095] 通过该步骤得到的所有第二级聚类候选中心,记做  $SC = \{g_1, \dots, g_F\}$ , 其中  $g_f \in \{r_1, \dots, r_N\}$ , 其中  $f \in \{1, \dots, F\}$ ,  $F = V \times N / 100$ ;

[0096] 3) 使用局部特征改进

[0097] 上述所有步骤都考虑的是图像的全局方面的特征,对于场景区分,建筑物识别等局部特征一直都体现了其有效性。因此,本发明在这一步做了局部特征确认,旨在保证选出来的图像组能够更好地表示输入图像,以达到估计准确的目的。在局部特征的相似性度量中,使用了本发明使用了两种不同的方法,一种是基于 BoW 直方图的相似性度量,另外一种是基于视觉词汇倒排索引结构的相似性度量。比对的是图像与各个离线系统中选择出来的代表性图像组之间的相似度,期待通过选择相似的图像组,并使用图像组的 GPS 位置信息来估计图像的拍摄地点位置。

[0098] 基于 BoW 直方图的相似性度量

[0099] 在离线系统中,生产码数之后,我们对每个代表性图像组用 BoW 直方图进行描述。在此,对输入图像也生成相对应维数的 BoW 直方图,并表示为  $h(k)$ ,  $k = 1, \dots, Q$ 。本发明中使用四种常见的直方图距离度量方式进行距离度量,包括余弦距离(记为 COS),城市距离(MAD),欧式距离(MSD)和直方图相交(HIST)。其距离计算方法分别如下,

$$[0100] COS(i, j) = \frac{\sum_{k=1}^Q NH_{i,j}(k) \times h(k)}{\sqrt{\left[ \sum_{k=1}^Q (NH_{i,j}(k))^2 \right] * \left[ \sum_{k=1}^Q (h(k))^2 \right]}} \quad (6)$$

$$[0101] MAD(i, j) = \sum_{k=1}^Q |NH_{i,j}(k) - h(k)| \quad (7)$$

$$[0102] MSD(i, j) = \sum_{k=1}^Q (NH_{i,j}(k) - h(k))^2 \quad (8)$$

$$[0103] HIST(i, j) = \sum_{k=1}^Q \min(NH_{i,j}(k), h(k)) \quad (9)$$

[0104] 基于倒排索引结构的相似性度量

[0105] 倒排索引结构近些年来经常用于文本检索中,随着图像检索中 BoW 的不断发展,倒排索引也得到了广泛的应用。本发明中使用了基本的 TF-IDF 计算权值的方法。对每个图像组计算其相应的得分。得分计算方法如下,

$$[0106] Score(L) = \sum_{x=1}^Q \frac{\omega_x * Freq_L(x)}{Number_L * Frequency_x} \quad (10)$$

[0107] 其中  $Freq_L(x)$  是第  $x$  号词汇的频数,  $Number_L$  是图像中的所有视觉词汇的数目,  $Frequency_x$  是词汇在整个图像库中出现的次数,  $\omega_x$  表示的是对于输入图像词汇的权重。权重计算方法如下,

$$[0108] \omega_x = \frac{Freq_{input}(x)}{Number_{input}} \quad (11)$$

[0109] 其中  $Freq_{input}(x)$  表示 BoW#x 的频次 and  $Number_{input}$  输入图像所包含的视觉词汇的总数。

[0110] 以上使用 BoW 直方图或者倒排索引结构对选取的第二级聚类候选中心进行确认后;再根据确认得到的结果使用 KNN 的方法对输入图像的地理位置进行推荐。

[0111] 实验结果及分析

[0112] 为了验证提出系统的性能,本文对比了 IM2GPS,田等人的空间编码的方法(记为 SC,包括 1-NN 和 K-NN 两种),以及李等人的基于 SVM 的地标建筑物的分类方法(记为 LC)。

[0113] 性能评估包括两个部分,第一部分是 GOLD 库与 GOLDEN 库上的交叉检验,第二部分是每个图像库内部的图像 GPS 位置估计的准确率和时间性能统计。

[0114] 评价准则

[0115] 1) 误判率

[0116] 在交叉检验中,如果 GOLD 库中的图像作为输入图, GOLDEN 库作为离线图像库的话,根据本文设定的准则应该是判断为不可估计,但是如果系统误判为图像可以 GPS 位置估计并且估计出了错误的 GPS 位置,那么就认为是误判的。反之, GOLDEN 库中的图像做输入图像,GOLD 做离线库的时候,原理也是相同的。本文中,ER 表示误判率,用 FN 代表发生误

判的图像数目, TN 表示所有用于交叉检验的图像总数。误判率的计算公式如下 :

$$[0117] ER = \frac{FN}{TN} \times 100\% \quad (3-12)$$

[0118] 2) GPS 估计的准确率与时间性能对比

[0119] 系统的准确率是通过从离线图像库中选取图像作为输入, 利用系统进行 GPS 位置估计, 然后判断与真实的位置是否一致进行准确率度量的。计算公式如下 :

$$[0120] AR = \frac{1}{G} \sum_{i=1}^G A_i \quad (3-13)$$

[0121] 式中 :

[0122] AR——整个图像库的图像 GPS 估计准确率 ;

[0123]  $A_i$ ——第 i 个景点的估计准确率 ;

[0124] G——所有测试的地理位置个数。

[0125] 针对每个景点分布计算准确的公式如下 :

$$[0126] A_i = \frac{NC_i}{NA_i} \times 100\%, i \in \{1, \dots, G\} \quad (3-14)$$

[0127] 式中 :

[0128]  $NC_i$ ——正确估计的图像个数 ;

[0129]  $NA_i$ ——该地理位置出所有测试图像的总数。

[0130] 与其他方法的对比

[0131] 为了公平起见, 所有的对比方法中, 使用的都只有图像内容。对于方法 SC, 1-NN 和 K-NN 都进行了对比, 其中 K 选取了是其性能达到最佳的值 :120。LC 方法中视觉词汇的规模为 6 万, 对比中, 提出的系统的基本设置参数为 R=32, M = 10, K=50, V = 100, 同时, 视觉词汇的规模为 6 万。基于直方图的相似性度量以及基于倒排索引结构的方法都参与了对比。实验结果如下表 3-1 和表 3-2 所示。

[0132] 表 3-1SC(1-NN), SC(K-NN), IM2GPS, LC 以及本文方法在图像库 COREL5000, OxBuil5000 和 GOLD 上的 GPS 估计准确度 (%)

	Data-set	SC	SC	IM2	LC	COS	MA	MS	HIS	IFS
		(1-NN)	(K-NN)	GPS		D	D	T		
[0133]	<b>COREL</b>				49.4	97.0	96.0	97.0	95.0	91.0
	<b>5000</b>	<b>58.64</b>	<b>76.01</b>	<b>45.98</b>	<b>3</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
	<b>OxBuil</b>				<b>53.9</b>	<b>91.0</b>	<b>90.0</b>	<b>90.0</b>	<b>89.0</b>	<b>87.0</b>
	<b>d5000</b>	<b>40.98</b>	<b>60.87</b>	<b>39.67</b>	<b>4</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
					<b>54.2</b>	<b>84.6</b>	<b>84.0</b>	<b>85.0</b>	<b>84.2</b>	<b>83.9</b>
	<b>GOLD</b>	<b>36.76</b>	<b>71.84</b>	<b>53.06</b>	<b>5</b>	<b>4</b>	<b>5</b>	<b>2</b>	<b>1</b>	<b>4</b>

[0134] 结果表明, 本文提出的系统性能明显优于其他方法, 不仅在 GOLD 库上, 同时也表现在其他两个测试库 OxBuil5000 和 COREL5000 上。三个测试数据集上 IM2GPS 的平均精度为 45. 98%, 39. 67%, 53. 06%。LC 的平均精度为 49. 43%, 53. 94%, 54. 25%。SC(K-NN) 在三个测试集的性能分别为 76. 01%, 60. 87% 和 71. 84%, 比其相应的 SC(1-NN) 性能有所提升。我们的方法根据 COS 进行举例度量方法在三个数据集上的结果分别为 97%, 91% 和

84. 64%。MAD, MSD, HIST, 和 IFS 下我们的方法比 IM2GPS, LC 和 SC 更好。

[0135] 表 3-2. SC(1-NN), SC(K-NN), IM2GPS, LC 以及本文方法在图像库 COREL5000, OxBuil5000 和 GOLD 上的 GPS 估计时间消耗 (ms)

[0136]	Data-set	SC (1-NN)	SC (K-NN)	IM2 GPS	LC	COS	MA	MSD	HIS	IFS
				GPS		D		T		
<b>COREL</b>										
	<b>5000</b>	<b>7.30</b>	<b>7.94</b>	<b>60.46</b>	<b>1.04</b>	<b>0.76</b>	<b>0.71</b>	<b>0.82</b>	<b>1.08</b>	<b>0.07</b>
<b>OxBuil</b>										
[0137]	<b>d5000</b>	<b>5.51</b>	<b>5.42</b>	<b>33.74</b>	<b>1.34</b>	<b>0.47</b>	<b>0.41</b>	<b>0.50</b>	<b>0.49</b>	<b>0.09</b>
	<b>GOLD</b>	<b>39.60</b>	<b>47.00</b>	<b>64927</b>	<b>2.89</b>	<b>0.96</b>	<b>0.93</b>	<b>1.03</b>	<b>0.99</b>	<b>0.16</b>

[0138] 通过实验发现,有效的结合全局特征和局部特征有利于图像 GPS 位置估计。因为 IM2GPS 仅仅利用了全局特征,其 AR 是比较低的。虽然 SC 利用局部特征,确忽略了全局特征所能够提供的有效信息。因此,我们的方法体现出了更好地性能。LC 的估计准确率较低的原因有两个。其中之一是,他们使用的 BoW 直方图的时候,局部特征的空间信息是完全被忽略的。另外一方面,是 SVM 分类受到训练集的影响是很大的,而一般情况下,不能保证训练集中的图像不包含噪声图像。

[0139] 在三个测试集上, IM2GPS 平均计算成本是 60.46ms, 33.74ms, 64927ms, 而 SC(K-NN) 是 7.30ms, 5.51ms, 39.60ms, LC 则分别是 1.04ms, 1.34ms, 2.89ms。COS, MAD, MSD, HIST, 和 IFS 计算成本都要低于 SC, LC, 和 IM2GPS。对于大规模数据集, IFS 的有效性体现的非常明显,平均计算成本是 0.117ms, 这只是 IM2GPS 消耗时间的  $1.8 \times 10^{-6}\%$ , 0.25% 的 SC(K-NN), 12.19% 的 COS, 12.58% 的 MAD, 11.36% 的 MSD, 以及 11.82% 的 HIST 计算成本。

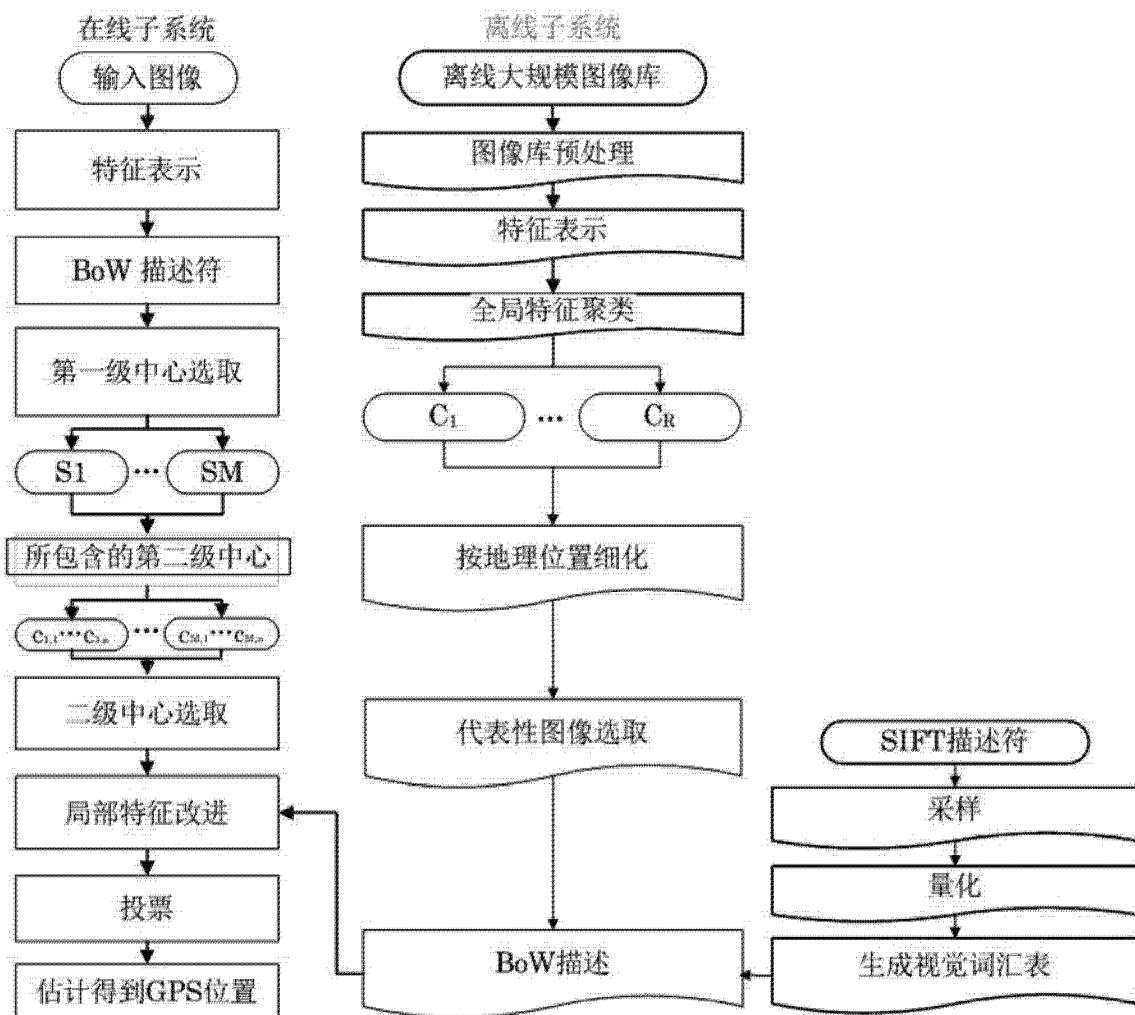


图 1

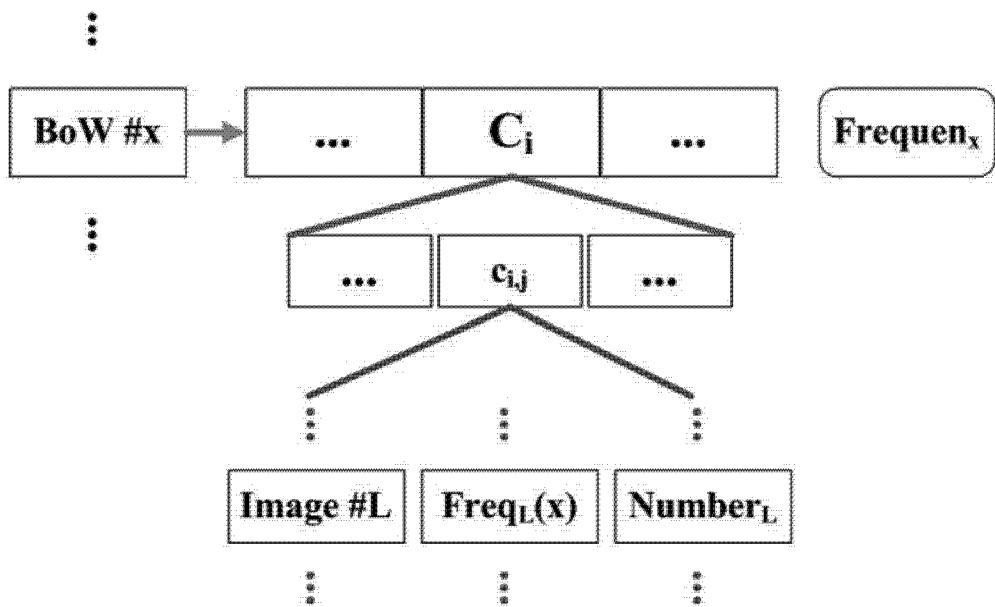


图 2