

文章编号:1001-9081(2014)04-1172-05

doi:10.11772/j.issn.1001-9081.2014.04.1172

## 基于加权特征空间信息视觉词典的图像检索模型

董 健\*

(盐城师范学院 信息科学与技术学院, 江苏 盐城 224002)

(\* 通信作者电子邮箱 [dongjyeah@yeah.net](mailto:dongjyeah@yeah.net))

**摘要:**针对传统的视觉词袋模型中视觉词典对底层特征量化时容易引入量化误差,以及视觉单词的适用性不足等问题,提出了基于加权特征空间信息视觉词典的图像检索模型。从产生视觉词典的常用聚类算法入手,分析和探讨了聚类算法的特点,考虑聚类过程中特征空间的特征分布统计信息,通过实验对不同的加权方式进行对比,得出效果较好的均值加权方案,据此对视觉单词的重要程度加权,提高视觉词典的描述能力。对比实验表明,在 ImageNet 图像数据集上,相对于同源视觉词典,非同源视觉词典对视觉空间的划分影响较小,且基于加权特征空间信息视觉词典在大数据集上更加有效。

**关键词:**图像搜索;视觉词袋模型;加权特征空间信息;视觉词典;聚类算法

**中图分类号:** TP391    **文献标志码:**A

## Image retrieval model based on visual vocabulary with weighted feature space information

DONG Jian\*

(College of Information Science and Technology, YanCheng Teachers University, YanCheng JiangSu 224002, China)

**Abstract:** Concerning the quantization error when the local features were quantified by the visual vocabulary in traditional Bag-of-Visual-Word (BoVW) model, an image retrieval model based on visual vocabulary with weighted feature space information was proposed. Considered the clustering method which was used to generate the visual codebook, the statistic information of the feature space was analyzed during the clustering process. Through the comparison of different weighting methods by experiments, the best weighting method, mean weighted average, was found to weight the visual words to improve the descriptive ability of the codebook. The experiment on ImageNet dataset shows that, compared to homologous visual codebook, non-homologous visual codebook has less impact on dividing the visual space, and the effects of the weighted feature space based visual codebook on big dataset are better.

**Key words:** image retrieval; bag of visual word model; weighted feature space; visual vocabulary; cluster algorithm

## 0 引言

智能终端设备如智能手机、平板电脑,以及数码相机和数码摄像机的普及和便携化,使这些多媒体采集设备已经成为人们日常生活中不可缺少的一部分。同时智能终端与互联网之间的交互越来越方便,人们大量地使用这些设备来上传、浏览和交流拍摄的数字图像。据 Intel 统计<sup>[1]</sup>,仅在 Flickr 上就有2000万的图像浏览次数。据统计<sup>[2]</sup>,截至 2011 年 8 月, Flickr 上的图像总数已达 60 亿张。相对于文字的信息而言,数字图像具有直观易于理解、信息层次感强、重点突出的特点。面对这无比庞大的图像数据库,人们自然会产生这样的需要:从这海量图像中寻找满足自己要求的图像。因此,人们开发了各种各样的图像检索系统,以满足不同人群的各式各样的需求。

基于内容的图像检索 (Content Based Image Retrieval, CBIR) 模型<sup>[3-5]</sup>的核心思想是模拟人对图像的认知,通过对图像的内容进行计算机层面的分析,根据内容建立语义索引,记录图像深入层面的信息,从而更好地理解用户的需求,呈现更高质量的检索效果。目前有多个在线的基于内容的图像检索系统,比如国外的 TinEye<sup>[6]</sup>, Google Images<sup>[7]</sup> 和百度的

识图等。

近年来,国内外对图像检索的研究主要集中在对图像的底层视觉特征以及全局特征的相关反馈的研究。Haas 等<sup>[8]</sup>提出了基于感知学习的相关反馈算法;Wang 等<sup>[9]</sup>提出了带重启的随机游走的图像注释提前算法,从其注释中找到相关视觉特征;针对图像旋转对图像检索造成的影响,朱正礼等<sup>[10]</sup>提出了预计非下采样轮廓变换、灰度共生矩阵和新相似性度量的旋转不变纹理图像检索算法。受文本领域信息检索的启发,Sivic 等<sup>[11]</sup>将文本信息检索中的词袋模型 (Bag of Word, BoW) 引入了基于内容的图像检索领域,提出了视觉词袋模型 (Bag of Visual Word, BoVW),其核心是将图像局部特征量化到事先聚类的视觉词典上,然后统计一幅图像中出现的视觉单词数量,将图像表示成视觉单词的统计直方图,然后通过倒排索引建立图像集合的数据库表示。尽管 BoVW 模型是一个较好的用于表示海量图像的方法,但是它仍然有很多不足之处,主要体现在视觉词典的使用上。常用的视觉词典主要具有以下几个缺点:1)在使用视觉词典量化图像的视觉单词时引入了噪声;2)视觉词典中单词的“多义性”没有得到充分挖掘;3)视觉词典这种表示形式本质上丧失了局部特征的空间信息;4)视觉单词和视觉词典的描述能力没有得到充

收稿日期:2013-10-12;修回日期:2014-01-10。    基金项目:盐城师范学院自然科学研究基金资助项目(11YCKL032)。

作者简介:董健(1980-),男,江苏盐城人,讲师,硕士,主要研究方向:多媒体信息处理、计算机图形图像处理。

分的挖掘。

针对上述问题,本文从用于产生视觉词典的常用聚类算法入手,分析和探讨了这些聚类算法的特点,同时考虑聚类过程中利用特征空间中特征分布的统计信息,产生视觉词典的优劣描述,并充分利用聚类算法产生的聚类中心会向特征空间中特征密集的方向移动这一特性,来提高视觉词典的表示能力。该方法与传统的词袋模型和词频-逆文档频率(TF-IDF)加权方式紧密结合,可以提升视觉词典的可迁移性和图像检索的检索效果。

## 1 BoVW 图像检索模型

### 1.1 基本框架

检索系统的目标是:针对用户输入的查询图像,系统从图像数据集中查找并返回最为相似的图像。一般来说,整个检索系统可以粗略地划分为两个部分:图像数据集合的离线索引系统和针对图像的在线检索系统。BoVW 图像检索模型的框架如图 1 所示。

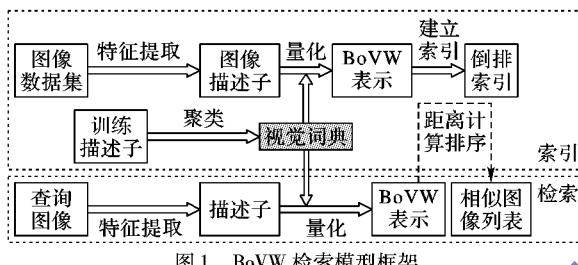


图 1 BoVW 检索模型框架

离线索引过程的相关步骤如下:

- 1) 从图像中提取底层的局部特征描述;
- 2) 选择部分训练图像,对其局部特征进行聚类得到视觉词典;
- 3) 利用该词典对数据集图像的特征进行量化,统计每幅图像量化后的相同视觉单词数目,得到图像的视觉单词直方图表示(BoVW 表示);
- 4) 为图像集合建立基于视觉词典的倒排索引。

倒排索引指的不是按照“图像-视觉单词直方图”的组织形式存储图像的 BoVW 表示,而是通过“视觉单词-出现该视觉单词的图像”这种方式来组织 BoVW 表示。在建立倒排索引时,为了区分不同视觉单词的重要性,需要统计每个视觉单词在整个图像中出现的次数,出现次数多的视觉单词重要程度相应地就低一些。加上之前的视觉单词数目统计,这构成了图像检索领域词袋模型的“词频-逆文档频率”(TF-IDF)权重度量方法。

在线检索过程的相关步骤如下:

- 1) 对查询图像提取相同的局部特征;
- 2) 利用相同的视觉词典对查询图像的特征进行量化;
- 3) 产生类似的视觉单词直方图表示;
- 4) 利用倒排索引结构,计算查询图像的 BoVW 表示和倒排索引中图像集合 BoVW 表示的向量距离,然后输出距离最小,即相似度最高的图像列表,显示在屏幕上,即为检索结果。

### 1.2 对视觉词典的分析

从 1.1 节中可以看出,视觉词典用于将图像转化为图像的 BoVW 表示,因此视觉词典的质量直接影响着图像检索的效果。因此如何产生视觉词典对于视觉词袋模型来说至关重要。经过大量文献中的实验验证<sup>[12-14]</sup>,视觉单词的数量需要

达到百万量级,才能取得最佳的检索效果。然而,产生这个数量量级的视觉词典需要使用经过优化的聚类算法。常用的视觉词典生成方法是进行分层  $k$  均值聚类(Hierarchical  $k$ -Means Clustering)<sup>[15]</sup>,从而产生所谓的词表树(Vocabulary Tree)<sup>[15]</sup>。该方法的思想在于将计算复杂度极高的大数量  $k$  均值聚类分成多层的树结构形式进行:

- 1) 对待聚类数据进行一个较小的  $k$  均值聚类计算,产生  $k$  个聚类中心,从而将数据空间划分成  $k$  个簇;
- 2) 在每个簇中依次进行  $k$  均值聚类,继而产生了  $k$  份  $k$  个较小的聚类结果;
- 3) 以此类推,进行  $h$  次这样的聚类后,即可得到一个  $k^h$  个叶子节点,高度为  $h$  的完全  $k$  叉树。

这个  $k$  叉树的所有节点都是聚类中心,可以用来产生视觉词典。在特征量化时,通过得到的语义树子树的根部向下,寻找每个特征在整个语义树中距离最近的视觉单词。

该聚类方法存在一定缺陷:在分层聚类的过程中,聚类中心将逐渐向特征空间中特征密度较大的地方偏移。由于这一因素,导致聚类结果往往不甚理想。

## 2 基于加权特征空间信息的视觉词典

### 2.1 对特征空间的分析

特征空间指的是由特征提取(Feature Extraction)步骤得到的特征向量(Feature Vector)生成的向量空间,该概念与线性代数中对应于特征向量(Eigen Vectors)的特征空间不一样。特征空间的每个点均对应一个特征向量,且每个定长的特征向量均对应于等维度的特征空间中的一个点。对特征空间和空间中的关系更简洁与准确的描述,意味着对由特征向量表示的图像数据集更为紧致与精确地表达,也就是说能够提高图像检索系统的图像存储量与检索的效果。因此,如何刻画特征空间,如何用一定的方式来描述在特征空间中匹配的特征向量之间的关系,是非常重要的问题。

视觉词典本质上是一种对特征空间中向量匹配关系的描述。一般来说,通过特征提取得到的特征向量,利用视觉词典量化为特定的视觉单词。这个过程在特征空间中的表现即为对特征空间中点的聚类。因此,基本的视觉词典模型中视觉词典对特征空间中点与点之间关系的刻画方法,即对距离较近、成簇的点进行聚类。如图 2 中,圆圈代表视觉单词,以这些视觉单词为中心,将特征空间划分为多个子空间,也称为簇。子空间中的单个特征,量化为子空间中的某个视觉单词。这种认为被分到一个簇中的两个点即匹配的量化方法创建了一种较为粗糙的特征空间中点的匹配关系。

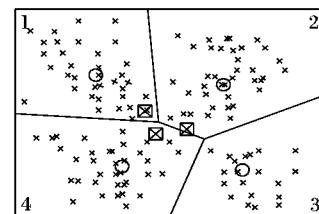


图 2 视觉词典对特征空间进行划分

区域 2 中的方块中的点与四个视觉单词的距离都比较近,从图中可看出该点与其他点形成的较为密集的团的距离都比较远,本不属于任何一个簇,但是在该视觉词典中被量化到区域 2 中。本来应该与区域 2 中方块的点距离较近的两个

方块中的点在该视觉词典的量化方法下被分到了不同的视觉单词产生的簇中, 这显然不够合理, 而这一现象产生的根本原因在于对视觉单词的聚类过程中仅仅考虑到该点与聚类中心的距离, 而没有考虑到视觉词典中视觉单词的重要性的不同, 也就是视觉单词的权重问题。因此, 针对该缺点, 本文提出了基于加权的特征空间信息的视觉词典的方法。

## 2.2 基于加权特征空间信息的视觉词典

本节分析了视觉词典对特征空间中点与点间关系描述能力的重要性, 并提出一种利用特征空间统计信息来对视觉词典中视觉单词的重要程度进行度量的方法, 来提高视觉词典对点间关系的描述能力, 从而提高检索的效果。选用对检索效果的评价标准平均正确率均值 (mean Average Precision, mAP) 来评估视觉词典的可迁移性。

在传统的视觉词典方法中, 单张图像中的特征量化为视觉词典中最近邻的视觉单词, 图像表示为一个直方图向量  $X$ , 其中  $X \in i^k$ ,  $k$  为视觉词典词汇量的大小。本文方法受到文献 [16–17] 的启发, 这两篇文章均指出: 随着  $k$  均值聚类的聚类中心数目的增加, 聚类中心会向特征空间中特征点更为密集的方向移动。那么, 如果一个簇中特征分布得越“松散”, 那么这个簇对应的视觉单词的区分力可能更强。为了研究特征子空间的簇中特征点分布与视觉词典的关系, 统计了每个簇中特征点到聚类中心的距离的分布信息随着图像数据集和视觉词典词汇量的改变对图像检索效果产生的影响, 具体来说就是利用各种特征子空间簇的统计信息作为簇对应的视觉单词的权重, 对视觉词典进行加权: 计算统计向量  $D \in i^k$ , 然后每个视觉单词均被统计信息加权:

$$X_j^T = D_j \times X_j \quad (1)$$

其中:  $X_j^T$ 、 $D_j$  和  $X_j$  是向量  $X_i^T$ 、 $D$  和  $X_i$  的第  $j$  维上的值,  $X_i$  是图像集合的第  $j$  个视觉单词直方图, 即第  $j$  个图像的 BoVW 表示;  $X_j^T$  是加权后的 BoVW 表示, 代替  $X_i$  来描述图像。本文使用的方法是一种新的距离度量方法, 改变了簇中点到点的距离度量, 而并未对距离度量方法进行改变。

通过选择几种不同的加权策略, 来决定如何利用分布信息来度量视觉单词的重要程度, 如图 3 所示, 本文尝试了 5 种不同的特征信息分布加权向量  $D$  的方法。令  $d_{ij}$  为第  $j$  个簇中第  $i$  个特征点到该簇的聚类中心的距离, 其中  $1 \leq i \leq M_j$ ,  $1 \leq j \leq N$ ,  $M_j$  是第  $j$  个簇中特征点的数量,  $N$  是视觉词典中视觉单词的数量, 可以产生以下几种关于距离的函数:

$$m_j = \frac{1}{M_j} \sum_{i=1}^{M_j} d_{ij} \quad (2)$$

$$v_j = \frac{1}{M_j} \sum_{i=1}^{M_j} (d_{ij} - m_j)^2 \quad (3)$$

实际上,  $m_j$  和  $v_j$  即为每个簇中特征点到聚类中心距离的均值与平均方差。在这里使用均值向量  $\mathbf{m} \in i_+^N$  和平均方差向量  $\mathbf{v} \in i_+^N$  的如下形式作为不同的权重重量函数  $W(\mathbf{D})$ : 原始的向量、指数函数的形式与倒数, 即  $\mathbf{m}$ 、 $\mathbf{v}$ 、 $\exp(\mathbf{m})$ 、 $\exp(\mathbf{v})$ 、 $\exp(-\mathbf{m})$ 、 $\exp(-\mathbf{v})$  和  $\mathbf{m}^{-1}$ 。在图 3 中显示了其中的五种加权方式的视觉词典应用于检索的效果:  $\exp(\mathbf{m})$ 、 $\exp(\mathbf{v})$ 、 $\exp(-\mathbf{m})$ 、 $\exp(-\mathbf{v})$  和  $\mathbf{m}^{-1}$ 。在这些实验中, 混杂数据集大小包含  $1 \times 10^4$  张图像。在图 3 中: Same-w 和 Diff-w 曲线表示加权后的视觉词典应用于检索的效果, Same 和 Diff 曲线表示

未加权的视觉词典应用于检索; Same-w 和 Same 曲线代表同源视觉词典应用于检索, Diff-w 和 Diff 曲线代表非同源视觉词典应用于检索。

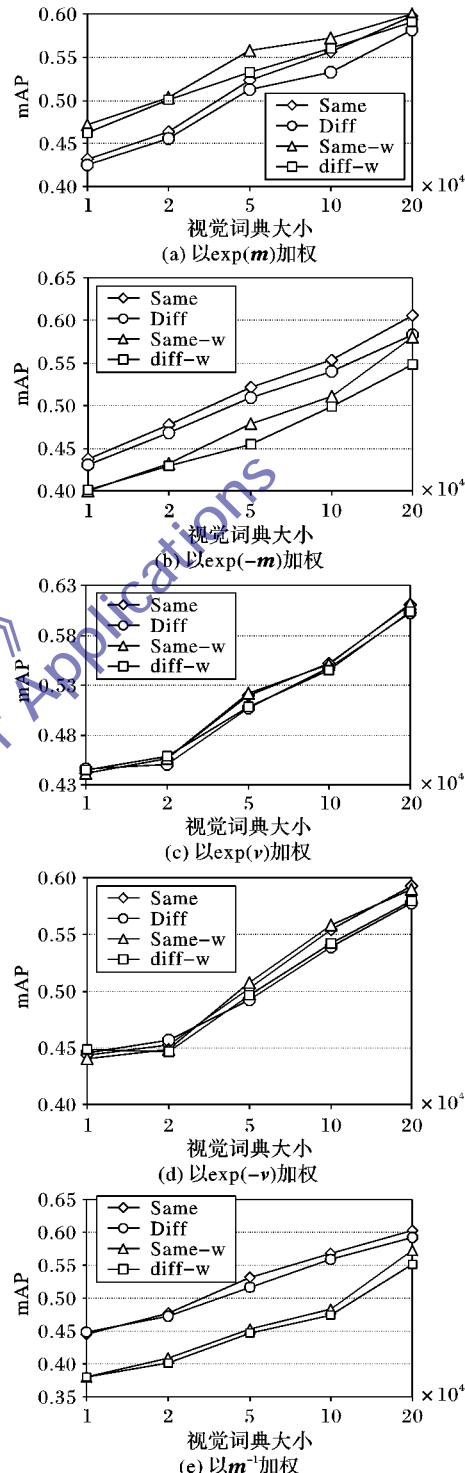


图 3 同源与非同源视觉词典不同加权对比

从图 3 中可以得出如下结论:

1) 随着视觉词典中视觉单词数量的增加, 检索的效果基本呈现越来越好的趋势, 这点和图像检索本身的特点相关, 视觉词典中视觉单词数量越多, 则该视觉词典更能体现图像的具体特征, 因此检索的结果必然更加精确。

2) 平均方差加权的方法应用于检索的效果不明显, 这是因为特征点到聚类中心的离散程度较大, 从而平均方差差异

较小,因此使用平均方差并不能区分各视觉单词的重要程度。

3) 利用均值的单调增函数对视觉词典进行加权,应用于检索的效果比较好,均值越小,说明该视觉单词更加靠近视觉中心,此视觉单词更能描述图像的特征空间信息;用非同源的视觉词典应用加权方法时,性能比未加权同源视觉词典方法好,因为给视觉单词加权的目的是为了最大限度地消除视觉词典对底层特征量化时引入量化误差。

因此,为了减少计算复杂度,本文选择了原始的均值作为对视觉词典进行加权的方法。

### 3 实验结果验证与分析

#### 3.1 对特征空间分布的分析实验

本节通过实验来对特征空间的分布特点进行具体的量化表示。所使用的局部特征是 DoG 关注点检测方法与尺度不变特征转换 (Scale-Invariant Feature Transform, SIFT) 描述子,通过分层  $k$  均值聚类方法产生词汇量不同的视觉词典。使用

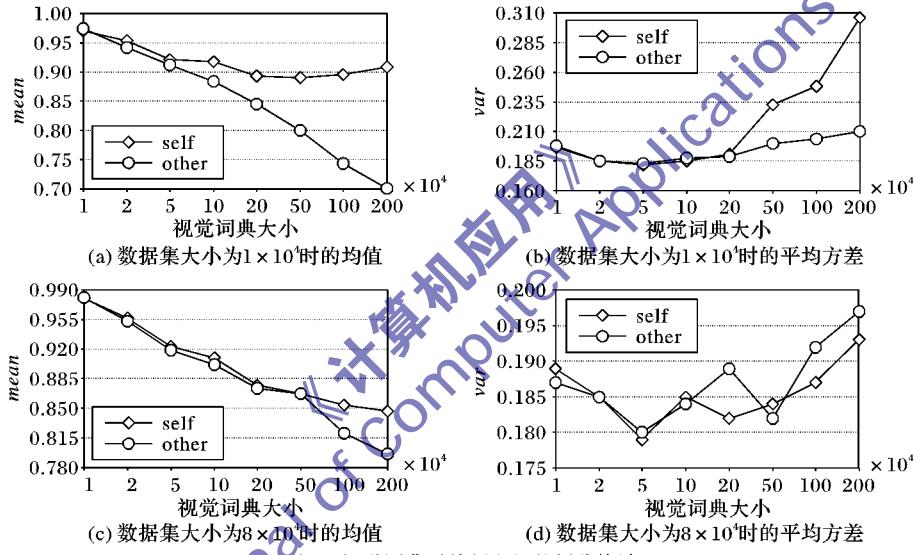


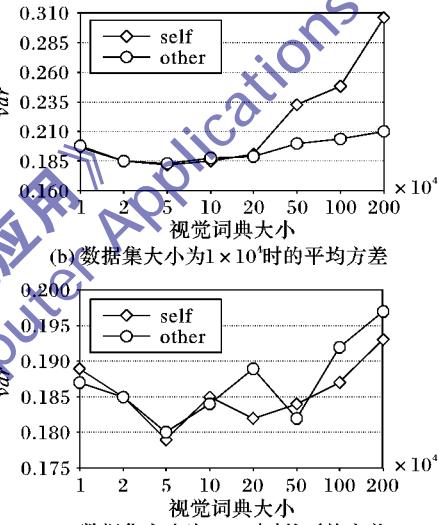
图4 视觉词典对特征空间的划分统计

从图4中可以得出如下结论:

1) 对于特征空间中点距离的均值来说,非同源视觉词典对特征空间进行划分,点距离的均值基本不会随着视觉词典中视觉单词数量的增加而变化,从图4(a)和(c)中可以看出,随着非同源视觉词典视觉单词数量由1万增加到200万,而均值由0.97下降到0.7,且呈直线下降;同源视觉词典对特征空间进行划分,点距离的均值随着视觉词典中视觉单词数量的增加而逐渐减小。之所以出现这种情况是因为非同源视觉词典只能对图像的整体特征空间信息进行划分,也就是说非同源视觉词典对图像的具体特征空间信息区分不明显,而对于同源视觉词典而言,视觉词典数量的增加更加反映了图像具体的视觉信息,用该视觉词典来划分视觉空间意味着更小的划分尺度,因此点距离的均值会逐渐减小。此外,随着视觉词典词汇量增大,同源视觉词典与非同源视觉词典的均值曲线的差距越来越大,如图4(a)和(c)中两条曲线的差距越来越大。这验证了随着 $k$ 均值聚类中心数目的增多,聚类中心会向特征空间中特征密集的方向移动,所以随着视觉单词数量的增加,聚类中心向密集区域移动的趋势也逐渐累积,从而均值逐渐减小。另外一种可能原因是,随着视觉词典的词

的词表树 (Vocabulary Tree) 的层数和分支数分别为:(10,4),(12,4),(9,5),(10,5),(12,5),(9,6),(10,6),(8,7),从而产生词汇数目约为1万、2万、5万、10万、20万、50万、100万和200万的视觉词典,并通过倒排表 (Inverted Index) 来索引图像数据集的BoVW表示。

本实验采用了ImageNet数据库<sup>[18]</sup>的子集:ILSVRC2011 (ImageNet Large Scale Visual Recognition Challenge 2011) 中的1000类数据作为干扰数据集 (Distractor)<sup>[19]</sup>。ImageNet图像数据集包含20000多类图像,按照WordNet中名词的组织层级形成树状结构,树上的每个节点对应某个类别的图像,一般为数百张。在2个不同级别的数据集上,选用多个不同视觉单词数量的视觉词典来统计视觉词典对特征空间的划分统计信息,所使用的度量标准为空间中点与聚类中心之间欧氏距离的均值 (mean) 和平均方差 (var),如图4所示, self 曲线表示使用同源视觉词典对特征空间的划分, other 表示使用非同源视觉词典对特征空间的划分。



汇量越来越大,聚类中心对特征空间的划分越来越密集,每个簇的大小也逐渐减小,因此均值逐渐减小。但这种可能性被非同源的视觉词典上对均值的统计实验排除,因为在非同源的视觉词典对特征空间的划分中,并未出现这种现象。

2) 对于特征空间中点距离的平均方差来说,非同源视觉词典对特征空间进行划分,点距离的平均方差值基本不随着视觉词典中视觉单词数量的增加而变化,从图4(b)和(d)中 other 曲线可以看出,视觉词典的单词数由1万增加到200万,其平均方差在0.18~0.21变动;同源视觉词典对特征空间进行划分,点距离的平均方差值随着视觉词典中视觉单词数量的增加而逐渐增大,从图4(b)和(d)中 self 曲线可以看出,平均方差在0.18~0.31变动。随着聚类中心向密集区域移动的趋势逐渐累积,在每个簇中,虽然聚类中心周围的特征向量到该聚类中心的平均距离在减小,在平均方差的变化过程中,也有例外出现,比如在图4(d)中,平均方差是由0.19下降到0.18,然后才逐渐增加到0.19。这是因为每个簇中仍然存在一部分特征向量,这些特征向量到聚类中心的距离仍然比较大。正是因为这些点的存在,导致了点到聚类中心平均距离的变化范围较大,从而增大了平均方差。这些点使得

该簇不能成为一个“好”的簇,因为该簇中不仅存在一些分布比较集中的特征点,还存在着一些比较零散的特征点,将这些零散的特征点量化为该簇的聚类中心实际上是有问题的。该分析在本节之后的实验中也得到了验证。

3) 结合 1) 和 2) 中的现象,可以发现这也从另外一个方面证实了视觉词典存在可迁移性。从图 4 中可以看出,均值曲线和平均方差值曲线比较平缓,没有特别明显的变化趋势,比如在数据集为  $8 \times 10^4$  时,均值的变化范围为 0.8~1,而平均方差更小,在 0.18~0.19。究其原因是由于同源的视觉词典由产生该视觉词典的图像数据集的局部特征聚类得到, $k$  均值聚类过程的特点反映于平均方差值曲线上,如图 4(b) 所示,平均方差随着视觉词典数量的增加,随着视觉词典中特征向量的增加,同源视觉词典的检索效果要明显好于非同源视觉词典的检索效果,这是由于数据集越大,图像的特征向量越分散。而非同源的视觉词典中视觉单词的产生过程与图像数据集的局部特征没有任何关系,因此,同源的视觉词典反映了产生该视觉词典的图像数据集的特征分布的特点,利用同源视觉词典进行检索的效果会比非同源视觉词典表现得更好一些。

4) 随着数据集中图像数量的增加,均值曲线、平均方差值曲线之间的差距在逐渐减小。这表明:随着数据集增大,特征向量数量增加,特征空间中每个簇中的点数量也随之增加,同源和非同源的视觉词典对特征空间的划分之间的差距也随之减小。

### 3.2 对加权视觉词典方法的验证

接下来将在不同大小的数据集上对加权视觉词典方法的有效性进行验证。在 1 万和 8 万这两个不同级别的数据集上进行检索的实验,结果如图 5 所示, self-w 与 other-w 表示加权视觉词典方法。

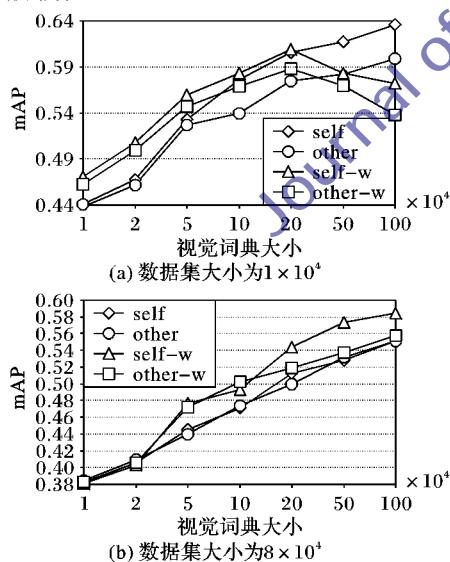


图 5 对加权视觉词典方法的验证

当实验所用的混杂数据集较小时,如 1 万张图像的情况,本文方法在视觉词典小于 20 万时表现得比未加权的结果好,但是超过了 20 万之后,加权的方法就不如未加权的方法,这是因为在数据集较小时,以均值为权重的视觉单词更容易体现出该特征在数据集中的特别之处,但是当数据集数量大于 20 万之后,这些加权后的视觉词典会“淹没”在大量的混杂数据集中。同时因为视觉词典过多带来的特征空间的过分割和

错误聚类,也在一定程度上降低了检索的效果:当实验数据集较小时,特征空间中特征点的数目也较少,因此产生过多的视觉单词将对特征空间进行过多的分割,导致每个簇中包含的特征数目较少,因此距离的均值不够稳定,用来作为权重的度量也不够好。当实验数据集的数目足够大时,每个簇中包含的特征数目上升,均值较为稳定了,因此权重的度量方法在这种情形下比没有权重的度量更好。

本方法在视觉词典上进行了加权,而图像索引系统为倒排表,由于倒排表表示的稀疏性,仅仅需要在图像的相似度计算时,对图像的词袋表示直方图中非 0 的 bin 进行加权。而从一幅中等分辨率的图像(如  $640 \times 480$  像素)中提取的 SIFT 描述子数量一般为几百到上千个,也就是说,对每幅图像量化后,得到的视觉单词种类一般也就几百种。从倒排表的结构来看,进行检索时,仅仅需要对查询图像中出现的视觉单词对应的列表进行检索。对于一个 10 万大小的图像数据集来说,粗略估计,倒排表中索引了约  $1000 \times 10$  万 = 1 亿条目,使用 100 万量级的视觉词典时,每个视觉单词对应的列表长度约为 100。也就是说,检索一张图像时,与标准的词袋模型方法相比,所增加的计算量为  $100 \times 1000 = 10$  万次浮点计算。这对于现代的 CPU 来说计算时间仅为毫秒量级。因此,本文的方法几乎不会降低词袋模型图像检索的速度。

### 4 结语

针对传统视觉词袋模型中对视觉单词的重要程度的估计不足,以至于在图像视觉特征量化时引起误差等问题,本文提出了一种基于特征空间的统计信息进行加权的视觉词典方法。对 BoVW 图像检索框架进行简要介绍,分析了视觉词典在该模型中的重要性;对特征空间中的特征分布进行分析,并根据前人所描述的聚类方法中聚类中心移动的特点,设计了具有针对性的基于加权特征空间信息的视觉词典,然后在多组图像数据集上进行检索实验,分析了不同加权方式的优劣,从中选择了最优的均值加权方法;最后利用统计实验和检索实验分析特征空间的分布以及验证加权视觉词典方法的有效性。实验结果表明,利用非同源视觉词典来划分特征空间,均值和平均方差的加权方法与视觉单词数量无关;而对于同源视觉词典而言,均值的加权会随着视觉单词数量增加而降低,而平均方差的加权会随着视觉单词数量增加而增加。最后,通过实验证明了本文方法简单而且有效,内存占用低,因此易于与其他对标准词袋模型方法结合。

### 参考文献:

- [1] TEMPLE K. What happens in an Internet minute? [EB/OL]. (2012-03-13) [2013-05-15]. <http://scoop.intel.com/what-happens-in-an-internet-minute/>.
- [2] How many photos are uploaded to Flickr every day, month, year? [EB/OL]. (2012-03-20) [2013-06-05]. <http://www.flickr.com/photos/franckmichel/6855169886/>.
- [3] SMEULDERS A W M, WORRING M, SANTINI S, et al. Content-based image retrieval at the end of the early years [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(12): 1349–1380.
- [4] LEW M S, SEBE N, DJERABA C, et al. Content-based multimedia information retrieval: state of the art and challenges [J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2011, 2(1): 1–19.

(下转第 1226 页)

几类嵌入式计算机，并且各自应用模式有所不同。应用结果证明，该采集系统在确保实时性的同时提高了系统的通用性和可扩展性，并且在计时精度上也有了数量级的提升，功能齐备，运行稳定，在各领域中需要进行时间信息采集的嵌入式系统内均可灵活应用。

#### 参考文献：

- [1] DANIELSON C F M, VANFRETTI L, ALMAS M S, et al. Analysis of communication network challenges for synchrophasor-based wide-area applications [C]// 2013 IREP Symposium-Bulk Power System Dynamics and Control-IX Optimization, Security and Control of the Emerging Power Grid. Piscataway: IEEE, 2013: 1–13.
- [2] AWEYA J, AL SINDI N. Role of time synchronization in power system automation and smart grids [C]// ICIT 2013: Proceedings of the 2013 IEEE International Conference on Industrial Technology. Piscataway: IEEE, 2013: 1392–1397.
- [3] CORREIA M, SOUSA J, COMBO A. Implementation of IEEE-1588 timing and synchronization for ATCA control and data acquisition systems [J]. *Fusion Engineering and Design*, 2012, 87(12): 2178–2181.
- [4] MAZUR D C, QUINT R D, CENTENO V A. Time synchronization of automation controllers for power applications [C]// Proceedings of the 2012 IEEE Industry Applications Society Annual Meeting. Piscataway: IEEE, 2012: 1–8.
- [5] HE H J, WANG Q Q, HE M X. The design of power quality on-line monitoring system based on GPS [C]// ICACC 2010: Proceedings of the 2nd International Conference on Advanced Computer Control. Piscataway: IEEE, 2010: 253–257.
- [6] REFAN M H, VALIZADEH H. Design and implementation of a GPS based DCS network time synchronization board [C]// CTPP 2011: Proceedings of the 3rd Conference on Thermal Power Plants. Piscataway: IEEE, 2011: 1–6.
- [7] ZHOU G P, QU S J, XU Y. The decoding device design of IRIG-B format time code [C]// CECNet 2012: Proceedings of the 2nd International Conference on Consumer Electronics, Communications and Networks. Piscataway: IEEE, 2012: 3575–3578.
- [8] ZHANG Z Y, WANG B, JIN E S. The design of the merging unit of real-time and synchronicity based on EP1C3T144 chip [C]// APAP 2011: Proceeding of the 2011 International Conference on Advanced Power System Automation and Protection. Piscataway: IEEE, 2011: 2341–2345.
- [9] MA H J, HU Y H, WU F P. Design and implementation of IRIG-B code time system using CPLD technique based on PCI bus [C]// Proceedings of the 2008 IEEE International Frequency Control Symposium. Piscataway: IEEE, 2008: 464–467.
- [10] FU Q Q, GAO Z Q, WU A P. The decoded design of IRIG-B [C]// ICECC 2011: Proceedings of the 2011 International Conference on Electronics, Communications and Control. Piscataway: IEEE, 2011: 2898–2901.
- [11] ZHOU G P, ZHANG J C. Research on high precision time signals transmission method [J]. *Sensor Letters*, 2013, 11(4): 722–725.
- [12] ZENG Y F, CHEN G J. Research on methods to improve precise time synchronization for IRIG-B code encoder [C]// APPEEC 2012: Proceedings of the 2012 Asia-Pacific Power and Energy Engineering Conference. Piscataway: IEEE, 2012: 1–4.

(上接第 1176 页)

- [5] DATTA R, JOSHI D, LI J, et al. Image retrieval: Ideas, influences, and trends of the new age [J]. *ACM Computing Surveys*, 2008, 40(2): Article No. 5.
- [6] TinEye. TinEye Reverse Image Search [EB/OL]. [2013-02-14]. <http://www.tineye.com/>.
- [7] Google. Google Images [EB/OL]. [2013-02-14] <http://images.google.com/>.
- [8] HAAS J P, HUESKE F, MARKL V. Detecting attribute dependencies from query feedback [C]// VLDB '07: Proceedings of the 33rd International Conference on Very Large Data Bases. Toronto: VLDB Endowment, 2007: 830–841.
- [9] WANG J, POHLMAYER E, HANNA B, et al. Brain state decoding for rapid image retrieval [C]// MM '09: Proceedings of the 17th ACM International Conference on Multimedia. New York: ACM, 2009: 945–954.
- [10] ZHU Z, ZHAO C, HOU Y, et al. Rotation-invariant texture image retrieval based on multi-feature [J]. *Journal of Nanjing University of Science and Technology: Natural Science Edition*, 2012, 36(3): 375–380. (朱正礼, 赵春霞, 侯迎坤, 等. 基于多特征的旋转不变纹理图像检索[J]. 南京理工大学学报: 自然科学版, 2012, 36(3): 375–380.)
- [11] SIVIC J, ZISSERMAN A. Video Google: a text retrieval approach to object matching in videos [C]// ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision. Washington, DC: IEEE Computer Society, 2003: 1470–1477.
- [12] PHILBIN J, CHUM O, ISARD M, et al. Lost in quantization: Improving particular object retrieval in large scale image databases [C]// CVPR 2008: Proceedings of the 2008 Computer Vision and
- [13] JEGOU H, DOUZE M, SCHMID C. Hamming embedding and weak geometric consistency for large scale image search [C]// ECCV 2008: Proceedings of the 10th European Conference on Computer Vision: Part I, LNCS 5302. Berlin: Springer-Verlag, 2008: 304–317.
- [14] JEGOU H, DOUZE M, SCHMID C. Improving bag-of-features for large scale image search [J]. *International Journal of Computer Vision*, 2013, 87(3): 316–336.
- [15] NISTER D, STEWENIUS H. Scalable recognition with a vocabulary tree [C]// CCVPR 2006: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2006: 2161–2168.
- [16] JURIE F, TRIGGS B. Creating efficient codebooks for visual recognition [C]// ICCV 2005: Proceedings of the 2005 Tenth IEEE International Conference on Computer Vision. Washington, DC: IEEE Computer Society, 2005: 604–610.
- [17] JI R, YAO H, XIE X, et al. Vocabulary hierarchy optimization and transfer for scalable image search [J]. *IEEE Multimedia*, 2012, 18(3): 66–77.
- [18] DENG J, DONG W, SOCHER R, et al. ImageNet: a large-scale hierarchical image database [C]// CVPR 2009: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2009: 248–255.
- [19] Image-Net Large Scale Visual Recognition Challenge 2011 (ILSVRC2011) [EB/OL]. [2013-04-23]. <http://www.image-net.org/challenges/LSVRC/2011/>.