# 非负矩阵分解耦合视觉词典的图像检索算法

李峰<sup>1</sup>, 应帅<sup>2</sup>, 卢文超<sup>2</sup>

(1.常州纺织服装职业技术学院 创意学院,常州 213164;2.吉安职业技术学院 机械与电子工程学院,吉安 343000)

摘要:目的 解决当前图像检索技术中,图像特征稀疏编码收敛速度慢,以及局部特征空间信息不足易 导致检索误差较大等问题,提出一种基于 l<sub>0</sub>稀疏约束非负矩阵分解耦合视觉词典优化的图像检索算法。 方法 首先,在非负矩阵分解(Non-negative Matrix Factorization, NMF)的基础上,对系数矩阵设置 l<sub>0</sub> 个约束来限制其稀疏性,从而定义一种 l<sub>0</sub>稀疏约束的 NMF 方法。再通过一种自适应序列词典初始化方 案,从训练样本获得词典的初始估计。然后,利用 l<sub>0</sub>稀疏约束的 NMF 来增强视觉词典,对图像局部描述符进行稀疏编码,并利用最大池化操作来生成聚合特征向量,从而保留局部描述符的关键属性。最后 根据得到的特征向量,引入 Minkowski 距离来衡量查询图像与数据库的相似性,输出检索图像。结果 实 验结果表明,与当前图像检索方案相比,所提算法具有更高的查准-查全率和收敛速度。结论 所提算法 返回的图像与查询图像相似度高,在包装商标检索等领域具有一定的参考价值。 关键词:图像检索;非负矩阵分解;视觉词典;稀疏编码;最大池化;Minkowski 距离 中图分类号:TP391.4 文献标识码:A 文章编号:1001-3563(2018)17-0215-08 DOI: 10.19554/j.cnki.1001-3563.2018.17.036

## Image Retrieval Algorithm Based on Non-negative Matrix Factorization Coupled Visual Dictionary

LI Feng<sup>1</sup>, YING Shuai<sup>2</sup>, LU Wen-chao<sup>2</sup>

(1.College for Creative Studies, Changzhou Vocational Institute of Textile and Garment, Changzhou 213164, China;
 2.School of mechanical and Electronic Engineering, Ji'an College, Ji'an 343000, China)

**ABSTRACT:** The work aims to solve such defects as the slow convergence speed of image feature sparse coding, and large retrieval error induced by insufficient local feature space information in current image retrieval technology. An image retrieval algorithm based on  $l_0$  sparse constraint non-negative matrix factorization coupled visual dictionary optimization was proposed. Firstly, on the basis of the non-negative matrix factorization (NMF),  $l_0$ -constraints were stetted on the coefficient matrix to limit its sparsity, so that a NMF framework for  $l_0$ -sparse constraints was defined. Then, an initialization scheme of adaptive sequence dictionary was proposed to obtain the initial estimation of the dictionary from the training samples. Then, the NMF with  $l_0$  sparse constraints was used to enhance the visual dictionary for sparse coding on image local descriptors, and the polymerization feature vectors were generated by the maximum pooling operation to retain the key attributes of the local descriptor. Finally, according to the obtained feature vectors, the Minkowski distance was introduced to measure the similarity between the query image and the database for outputting the retrieval image. The experimental results showed that the proposed algorithm had a higher precision-recall rate and faster convergence speed compared with the current image retrieval scheme. The image retrieved by the proposed algorithm has a high similarity to the query image, which has a certain reference value in the fields of package trademark retrieval, etc.

KEY WORDS: image retrieval; NMF; visual dictionary; sparse coding; maximum pooling; Minkowski distance

收稿日期: 2018-03-23

**基金项目:**江苏省自然科学基金(BK20151191)

作者简介:李峰(1980-),男,硕士,讲师,主要研究方向为图像检索、商标设计、信息处理。

随着商品品种的不断增加, 文字或图像上相似 的商标不断出现,导致商标识别检索的难度不断提 高<sup>[1-2]</sup>。图像检索在包装商标等领域具有重要的实际 应用意义。视觉词典(码本)是从一组预定义的函数 中创建的学习训练样本集,在实践中,学习词典在信 号和图像检索领域各种应用中具有较好的优越性[3-4]。 柯圣财等<sup>[5]</sup>设计了视觉词典优化与查询扩展的图像 检索方案,该方案通过引入密度聚类技术对尺度不变 特征变特征进行聚类,以构造视觉词典;通过图结构 的度量技术对检索结果排序,输出相关图像。由于该 技术采用的聚类算法易受到图像噪声的干扰,因此会 影响视觉词典的性能。BAO 等<sup>[6]</sup>通过引入 ROOTSift 算子来计算图像的特征值,并借助 K-means 聚类算 子构建动态随机视觉词典,以此建立单词直方图和倒 排索引序列,再将单词进行权值调整来提升匹配正确 性,最后引入 Euclidean 测量相似性,输出检索图像。 由于 ROOTSift 特征的特征维度较大,因此对于视觉 变化较小的相似图像,其检索效果不佳。此外, K-Means 对特征点聚类时,每次迭代都将数据点分配 给离之最近的聚类点。ZHU 等<sup>[7]</sup>通过特征值的能量布 局完成了对子空间的分解,再引入 Fisher 判别规则, 根据样本的空间信息得到最优投射的哈希方程,从而 形成特定的编码,以完成图像检索任务,最后引入随 机样本一致优化算子对初始返回的图像进行优化,提 高了检索性能。由于该技术没有考虑图像像素点的空 间关系,对局部特征利用不足,因此在对相似图像检 索时易出现误差。

针对上述问题,文中定义一种 ℓ<sub>0</sub>稀疏约束的 NMF 方案,并从训练样本获得词典的初始估计,降 低随机选取初始值对聚类效果的影响。再利用 ℓ<sub>0</sub>稀疏 约束的 NMF 来增强视觉词典,以加快收敛速度,减 小收敛误差。

## 1 非负矩阵分解优化

#### 1.1 非负矩阵分解 NMF

设*V*为从训练目标获取的*D*维局部描述符集合, *V* = [ $v_1$ , $v_2$ ,..., $v_N$ ] ∈  $R^{D \times N}$ 。通过对含有*K* 个基元的可 视化词典*W* 学习,表示为*W* = [ $w_1$ , $w_2$ ,..., $w_K$ ] ∈  $R^{D \times K}$ 。 NMF 将矩阵*V* 划分成 2 个非负矩阵*W* 和*H*的积。为 了估计*V* ≈ *WH* 的近似质量,需要建立一个适合的成 本函数。为此,文中引入了 Frobenius 范数<sup>[8]</sup>,其函 数如下:

$$C(\boldsymbol{W}, \boldsymbol{H}) = \left\| \boldsymbol{V} - \boldsymbol{W} \boldsymbol{H} \right\|_{F}^{2}$$
(1)

式中: *H* 为数据矩阵*V* 的系数向量集,表示为  $H = [h_1, h_2, ..., h_N] \in R^{K \times N}$ ; *F* 为矩阵的 Frobenius 范数。 为了学习码本, NMF 需要解决以下约束优化 问题:

$$\min_{\boldsymbol{W},\boldsymbol{H}} \operatorname{min}_{\boldsymbol{W},\boldsymbol{H}} C(\boldsymbol{W},\boldsymbol{H}) = \|\boldsymbol{V} - \boldsymbol{W}\boldsymbol{H}\|_{F}^{2}$$
subject to  $\boldsymbol{W},\boldsymbol{H} \ge 0$ 
(2)

在文献[9]中提出了一个迭代乘法更新方案,用 于求解(2)的局部极小值,定义如下:

$$\boldsymbol{H} \leftarrow \boldsymbol{H} \odot (\boldsymbol{W}^{\mathrm{T}} \boldsymbol{V}) / (\boldsymbol{W}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{H})$$
(3)

$$\boldsymbol{W} \leftarrow \boldsymbol{w} \odot (\boldsymbol{V}^{\mathrm{T}} \boldsymbol{H}) / (\boldsymbol{W} \boldsymbol{H} \boldsymbol{H}^{\mathrm{T}})$$

$$\tag{4}$$

式中: ①表示矩阵的乘积。

再通过引入交替非负最小二乘法<sup>[10]</sup>,通过反复执 行以下步骤来求解式(2)的最优问题,定义如下:

$$\boldsymbol{W}^{k+1} = \arg\min_{\boldsymbol{W} \ge 0} \left\| \boldsymbol{V} - \boldsymbol{W} \boldsymbol{H}^k \right\|_F^2$$
(5)

$$\boldsymbol{H}^{k+1} = \arg\min_{\boldsymbol{H} \ge 0} \left\| \boldsymbol{V} - \boldsymbol{W}^{k+1} \boldsymbol{H} \right\|_{F}^{2}$$
(6)

随后,通过梯度下降迭代更新准则来近似求解式 (2)的最优解,定义如下:

$$\boldsymbol{H}^{k+1} = \max[0, (\boldsymbol{H}^k - (\eta \cdot \nabla_{\mathrm{H}} C(\boldsymbol{W}^k, \boldsymbol{H}^k)))]$$
(7)

$$\boldsymbol{W}^{k+1} = \max[0, (\boldsymbol{W}^k - (\eta \cdot \nabla_{\mathbf{W}} C(\boldsymbol{W}^k, \boldsymbol{H}^{k+1})))] \qquad (8)$$

式中:  $\eta$  为预定义的步长;  $\nabla_{\mathrm{H}}C(W^{k}, H^{k})$ ,  $\nabla_{\mathrm{W}}C(W^{k}, H^{k+1})$ 分别为 W 和 H 的成本函数的梯度。

#### 1.2 稀疏编码

为了实现数据的降维,往往需要通过几种矩阵分 解技术对词典进行测量。其中,稀疏编码(Sparse Coding, SC)通过构造基元矩阵的过完备与系数矩阵 H列稀疏来实现这一目的<sup>[11]</sup>。这样,使用 L 个基元的 线性组成来近似表示局部描述符。通过 Frobenius 范 数可以用来量化数据矩阵 V 和实现模型之间的差异。 为了将较小的重建误差与稀疏性结合起来,可通过 SC 解决下列优化问题<sup>[12]</sup>:

$$\min_{\boldsymbol{W}, \boldsymbol{H}} \operatorname{mize} C(\boldsymbol{W}, \boldsymbol{H}) = \left\| \boldsymbol{V} - \boldsymbol{W} \boldsymbol{H} \right\|_{F}^{2}$$
subject to  $\chi(h_{i}), \forall h_{i} \in \boldsymbol{H}$ 

$$(9)$$

式中: $\chi(.)$ 为在 H 的列向量  $h_i$  上施加的稀疏函数。

在各种稀疏函数中,  $\ell_0 与 \ell_1$ 范数为最常见的分解 技术。前者计算给定向量的非零项的数目, 后者计算 向量中所有项的总和。对于一个固定词典 W, 求解带 有 $\ell_0$ 个约束是比较棘手的。对此, 往往采用贪婪方法 来近似求解<sup>[12]</sup>, 可以直接从一组训练样本中导出基向 量, 无需使用预先确定的词典来进行稀疏编码。

#### 1.3 基于 ℓ₀ 约束的 NMF

文中定义了一种稀疏的 NMF 学习方案,主要过 程为初始化得到非负 W 与 H,对 H 的每列进行稀疏 化,通过给定的稀疏度 s,计算  $\ell_0$ 范数,使其满足  $H = H \otimes (VW^T) \odot (WHH^T)$ 。对每一列元素设置  $\ell_0$ 个 稀疏约束来限制系数矩阵 H 的稀疏性, 表示如下:

$$\min_{\boldsymbol{W},\boldsymbol{H}} \operatorname{min}_{\boldsymbol{W},\boldsymbol{H}} C(\boldsymbol{W},\boldsymbol{H}) = \|\boldsymbol{V} - \boldsymbol{W}\boldsymbol{H}\|_{F}^{2}$$
subject to  $\boldsymbol{W},\boldsymbol{H} \ge 0$  (10)
$$\|\boldsymbol{h}_{i}\|_{0} \le L, \forall \boldsymbol{h}_{i} \in \boldsymbol{H}$$

稀疏约束通过利用 L 个非负基元来计算 V 的每 一列。如果利用 W 中的不同列元素来对某些高级特 征进行编码,那么每一个数据样本都可以复制不超 过 L 个这样的特征。通过保持 H 上的稀疏约束,同 样的原则也可以用来解决  $\ell_0$ 约束的 NMF。在算法 1 中给出了系数矩阵 H 上稀疏约束的 NMF 的过程。第 1 步是稀疏编码过程,其中非负稀疏编码被寻址。第 2 步是增强可视化词典 W。在此步骤中,H的非零项 也通过保持其稀疏结构而得到增强。详细的算法 1 如下:

输入:数据V,内部迭代数M1,外部迭代数M2

- 1. 内部迭代=1, 外部迭代=1
- 2. 随机初始化 W
- 3. 重复
- 4. H ← V 的非负稀疏码本
- 5. 重复
- 6. W ← 增强词典
- 7. H← 增强词典
- 8. 内部迭代=内部迭代+1
- 9. 直到内部迭代=M1
- 10. 外部迭代=外部迭代+1
- 11. 直到外部迭代=M2
- 12. 返回 W

## 2 视觉词典优化模型

#### 2.1 词典初始化

现有大多数 NMF 算法都是在词典矩阵 W的随机 非负初始化之后进行的。对字典完成初始化后, $\ell_0$ 稀 疏约束 NMF 必然收敛于局部极小值<sup>[13]</sup>。由此可见, 适当的非随机初始化可以加快收敛速度和收敛时的 最小逼近误差。为了填充视觉词典 W,所提出的自适 应顺序词典初始化算法每迭代 1 次,就可以从给定 的训练样本中识别 1 个基向量,因此学习码本的质 量对图像编码过程和后续的检索操作有着显著的影 响。首先,定义一个累积矩阵  $A^l = [a_1^l, a_2^l, ..., a_K^l]$ ,其 中, $a_k^l \in A^l$ ,为步长 l的第 k 个累积列向量。通过对 给定的 V 进行初始化累积,然后在每次迭代过程更 新累积矩阵。

通过对稀疏数据向量的迭代,然后将其正交投影 减去单个数据项,有助于从给定的观测值中识别出一 组线性无关的基向量。

#### 2.2 优化词典

完成对 W 的初始估计后,接下来对其进行优化。 文中提出的词典优化框架是一个迭代过程,涉及 2 个 主要步骤:数据矩阵 V 的非负稀疏编码和词典 W 的 增强。在每次迭代中,这 2 个步骤必须确保非负性以 及稀疏性约束。

#### 2.2.1 数据矩阵的非负稀疏编码

如前所述,稀疏编码的目的是通过视觉词典 W 中最基本元素的线性组合近似表示每个数据矢量,  $v_i \in V$ ,通常将其定义为以下形式的最小化问题:

$$\min_{\boldsymbol{H}} \min_{\boldsymbol{H}} \operatorname{mize} \|\boldsymbol{V} - \boldsymbol{V}\boldsymbol{H}\|_{F}^{2}$$
subject to  $\boldsymbol{H} \ge 0$ 

$$\|\boldsymbol{h}_{i}\|_{0} \le L, \forall \boldsymbol{h}_{i} \in \boldsymbol{H}$$

$$(11)$$

正交匹配追踪(Orthogonal Matching Pursuit, OMP)算子<sup>[14]</sup>是解决稀疏编码问题里最普遍、最简单的一种方案。这里基于 OMP 思想, 定义一个新的 OMP 用于非负稀疏编码中。该方案利用完全 L 基向量的非负稀疏码来逼近给定的数据矢量 $v_i$ 。为了得到式(11)的近似解,需要对数据矩阵 V的各个列重复执行。

#### 2.2.2 视觉词典增强

获得系数矩阵 H 后,下一步就是增强词典 W。 文中根据文献[15]的非负矩阵分解的梯度投影法来提 高视觉词典 W。所提出的 ℓ<sub>0</sub>约束梯度投影算法是一 种迭代过程,通过该过程,词典元素和相应的系数矩 阵可以交替地增强。系数矩阵 H 的增强是以只在非 零项保持不变的情况下更新非零项的方式来执行,矩 阵 W 和 H 根据它们的负梯度方向来更新,表示如下:

$$\boldsymbol{H}^{(k+1)} = \max[0, (\boldsymbol{H}_{\rm S}^{k} - (\alpha_{\rm H}^{k} \cdot \nabla_{\rm H} C(\boldsymbol{W}^{k}, \boldsymbol{H}^{k})))] \quad (12)$$
$$\boldsymbol{W}^{(k+1)} = \max[0, (\boldsymbol{W}_{\rm S}^{k} - (\alpha_{\rm H}^{k} \cdot \nabla_{\rm W} C(\boldsymbol{W}^{k}, \boldsymbol{H}^{k+1})))] \quad (13)$$

式中:  $\nabla_{\boldsymbol{H}}$  为偏导数;  $\nabla_{\boldsymbol{H}} C(\boldsymbol{W}^{k}, \boldsymbol{H}^{k})$ ,  $\nabla_{\boldsymbol{W}} C$ 

 $(\boldsymbol{W}^{k}, \boldsymbol{H}^{k+1})$ 分别表示 $\nabla_{\mathrm{H}} C(\boldsymbol{W}^{k}, \boldsymbol{H}^{k}) = \boldsymbol{W}^{(k)^{T}} (\boldsymbol{W}^{k} \boldsymbol{H}^{k} - \boldsymbol{V}),$  $\nabla_{\mathrm{W}} C(\boldsymbol{W}^{k}, \boldsymbol{H}^{k+1}) = (\boldsymbol{W}^{k} \boldsymbol{H}^{(k+1)} - \boldsymbol{V}) \boldsymbol{H}^{(k+1)^{T}}; \ \boldsymbol{\alpha}_{\mathrm{H}}^{k}, \ \boldsymbol{\alpha}_{\mathrm{W}}^{k} \mathcal{H}$ 別为 **H** 和 **W** 第 *k* 次迭代的学习率。

为了适应于公式(10)稀疏约束 NMF,首先, 通过柯西点来定义学习速率的初始估计:

$$\alpha_{\rm W}^0 = \frac{G_{\rm W}^1 G_{\rm W}}{G_{\rm W}^T Z_{\rm W} G_{\rm W}} \tag{14}$$

$$\alpha_{\rm H}^0 = \frac{G_{\rm H}^T G_{\rm H}}{G_{\rm H}^T Z_{\rm H} G_{\rm H}} \tag{15}$$

式中:  $G_H$ ,  $G_W$ 分别为在学习率上的梯度矩阵  $\nabla_{\mathrm{H}}C(W^k, H^k)$ ,  $\nabla_{\mathrm{W}}C(W^k, H^{k+1})$ 所对应的列向量;  $Z_{\mathrm{H}}$ ,  $Z_{\mathrm{W}}$ 分别为 H 与 W 的成本函数  $\|V - WH\|_2^F$  的 Hessians 方程, 表示如下:

$$Z_{\rm H} = \prod \otimes \boldsymbol{W}^{\rm T} \boldsymbol{W} \tag{16}$$

$$Z_{\rm W} = \boldsymbol{H}\boldsymbol{H}^{\rm T} \otimes \boldsymbol{\varPi} \tag{17}$$

式中: **Π** 为单位矩阵; ⊗为 Kronecker 乘积<sup>[16]</sup>。 使用柯西点估计的学习率有助于找到 **W** 和 **H** 梯 度上的值,同时最小化成本函数。其次,在每一次迭 代过程中,均采用基于 Armijo 线搜索规则来迭代更 新学习速率<sup>[17]</sup>。Armijo 条件保证学习率在这样一种 方式改变:在新的系数矩阵估计中,成本函数值应该 保持一个较小值。根据步长 α<sup>k+1</sup><sub>H</sub>, α<sup>k+1</sup><sub>W</sub>, Armijo 规 则满足以下 2 个条件:

$$C(\boldsymbol{W}^{k}, \boldsymbol{H}^{k+1}) - C(\boldsymbol{W}^{k}, \boldsymbol{H}^{k}) \leq (\boldsymbol{H}^{k+1} - \boldsymbol{H}^{k})$$
  

$$\nabla_{\mathrm{H}} C(\boldsymbol{W}^{k}, \boldsymbol{H}^{k})$$
(18)

$$C(\boldsymbol{W}^{k+1}, \boldsymbol{H}^{k}) - C(\boldsymbol{W}^{k+1}, \boldsymbol{H}^{k}) \leq (\boldsymbol{W}^{k+1} - \boldsymbol{W}^{k})$$
  
$$\nabla_{\mathbf{W}} C(\boldsymbol{W}^{k}, \boldsymbol{H}^{k})$$
(19)

通过对以下 2 个条件的评估,以确定每次迭代中 适当的学习率,定义如下:

由于没有相关限制,通过式(13)直接更新非负 词典 W。另一方面,在增强 H 时需要保持稀疏性。 为此,在稀疏编码完成后,保持在每个 h<sub>i</sub>中对应于非 零项的索引列表。这保证了在视觉词典 W 的细化过 程中系数矩阵 H 列的 ℓ<sub>0</sub>稀疏约束。在每一次迭代中, Armijo 规则确定的学习率可以有效减小成本函数值, 这有助于在词典优化过程中更快地收敛。

## 3 基于增强视觉词典的图像检索

近年来,研究提出了一种基于倒排文件的词汇树 索引局部图像特征,这有助于使用更大的词汇表以获 得更好的检索精度。由于特征量化过程中的信息丢失 和各种识别局部特征空间信息的缺乏,检索精度往往 受到一定程度的限制,因此,对初始匹配处理的图像 采用重新排序技术来提高精度,如随机抽样一致 (RANSAC)通常被用来提高精度<sup>[18]</sup>。在特征量化中, 将一组局部描述符转换成一个紧凑的固定尺寸特征 向量,通过少量的字节表示编码全局描述符的图像, 同时保留局部描述符继承的关键属性。编码过程将每 个 *D* 维局部图像描述符映射成另一种形式,而池化 过程是将所有编码特征聚合成单个向量。对此,文中 提出了新的检索算法,其过程见图 1。



图 1 文中算法流程 Fig.1 Flow chart of the proposed algorithm

#### 3.1 局部图像描述符的稀疏编码

所提出的方法以编码和基于池化的流水线为基础,构建了适合于检索任务的聚合图像特征向量。在编码阶段,利用自带 $\ell_0$ 约束的 NMF 来形成视觉词典,从而获取图像集的局部描述符对应的稀疏码。设 F 为从图像中提取的局部描述符 集,定义为 F =  $[f_1, f_2, ..., f_N] \in R^{D \times N}$ 。给出了一个带 K 个基向量的学习视觉词典 W,表示为 W =  $[w_1, w_2, ..., w_K] \in R^{D \times K}$ ,对于每一个  $f_i \in R^{D \times 1}$ 编码操作,计算其对应的向量 $c_i \in R^{K \times 1}$ 。随后,通过最小化约束优化问题得到系数矩阵 C =  $[c_1, c_2, ..., c_N] \in R^{N \times K}$ ,定义如下:

min imize 
$$\|\boldsymbol{F} - \boldsymbol{W}\boldsymbol{C}\|_{F}^{2}$$
  
subject to  $\boldsymbol{C} \ge 0$  (22)  
 $\|\boldsymbol{c}_{i}\|_{0} \le L, \forall \boldsymbol{c}_{i} \in \boldsymbol{C}$ 

式中:  $c_n$ 为第n个局部描述符编码,每列  $c_n \in R^{K \times l}$ 。算法3对每个局部图像特征 $f_i$ 迭代计算, 以形成最终的系数矩阵C。

#### 3.2 特征聚合

对于一个含有 N 个局部描述符  $F = [f_1, f_2, ..., f_N]_{D \times K}$  的图像,相应的稀疏编码  $C = [c_1, c_2, ..., c_N]_{K \times N}$ 。随后,将这些稀疏编码聚合到一个 K 维的特征向量  $P = [p_1, p_2, ..., p_K]_{K \times N}$ 中。通过池化操作,将其聚合局部稀疏编码的特征用于检索操作。对于一组 N 个稀疏编码描述符  $[c_1, c_2, ..., c_N]$ ,可以将最大池化操作定义为<sup>[19]</sup>:

$$p = \max(c_1, c_2, ..., c_N)$$
 (23)

#### 3.3 图像的相似性度量

在提取聚合特征向量之后,下一步是确定查询与

数据集中所有图像的对应关系。在文中算法中,采用 Minkowski距离的第1顺序来衡量查询图像与数据库 图像的相关性,定义如下:

$$D(P^{q}, P^{d}) = \sum_{i=1}^{K} \left| p_{i}^{q} - p_{i}^{d} \right|$$
 (24)

式中: q,d 分别表示查询图像与数据库图像;  $p_i^q, d_i^d$ 分别为聚集特征向量  $P^q, P^d$ 的第i个组件。

通过合并某些索引结构,可以加快搜索过程。聚 合图像特征的每个维度对应于所学的视觉词典中的 基向量,这将触发倒排文件结构<sup>[20]</sup>用于整理各种数据 集中的图像。为了计算图像相似性,通过在给定的查 询目标中选择具有非零条目的视觉项完成行交叉,从 整个集合中筛选出一组候选图像。然后只在这些图像 和查询之间计算相关性得分。

## 4 实验与分析

为验证提出算法的检索性能,选取常用的 Corel-1K数据库进行测试。Corel-1K主要通过10种 不同场景图像组成<sup>[21]</sup>:非洲、海滩、建筑、汽车、 恐龙、大象、花朵、马儿、山、食品,每种含100幅, 共1000幅,见图2。测试条件: Inteli7 4核CPU,



图 2 Corel-1K 图像示例 Fig.2 Image example of Corel-1K

3.50 GHz,8GB RAM,64 位 WIN8 系统,并通过 Matlab 2007 进行仿真分析。为了体现提出算法的优越性,将文献[5—7]作为参照组。实验参数设置:稀疏级 *L*=10, *W*=30\*50, *N*=1000, *K*=100。

为了衡量算法的检索性能,借助检索中最常用的 评价指标:查准率(Precision)、召回率(Recall)、精 度(Accuracy)度量。其中, Precision和Recall分别作 如下定义<sup>[22]</sup>:

$$P = \frac{T_{\rm p}}{T_{\rm p} + F_{\rm p}} \tag{25}$$

$$R = \frac{T_{\rm p}}{T_{\rm p} + F_{\rm N}} \tag{26}$$

式中:  $T_p$ 为返回正确的图像数量;  $F_p$ 为误检索数量;  $T_N$ 为漏检的数量。

精度 Ac 计算如下:

$$A_{\rm C} = \frac{T_{\rm p} + T_{\rm N}}{T_{\rm p} + F_{\rm p} + T_{\rm N} + F_{\rm N}}$$
(27)

式中: A<sub>N</sub>为真实图像被正确检索数量。

以 Corel -1K 数据集中的"食物"为查询对象,结果 见图 3。其中图 3a 为查询图像,图 3b-e 依次为文献 [5]、文献[6]、文献[7]以及文中方案返回的前 16 幅结 果。根据图 3 的检索结果可以看出, 文献[5]算法返回 的 16 幅图像中出现了 3 幅不相关图像(2 幅"非洲" 图像,1幅"公共汽车"图像);文献[6]出现了2幅不 相关图像(2幅"非洲"图像); 文献[7]出现了2幅不 相关图像(1幅"非洲"图像,1幅"公共汽车"图像); 文中算法返回了1幅不相关图像("非洲"图像)。依 据返回结果得知, 文中算法检索效果最优, 返回的对 象与查询目标相似度较高。文中算法取得了优异的检 索性能,主要是采用了 l<sub>0</sub>稀疏约束的 NMF 增强的视 觉词典,提高了算法效率,并通过视觉词典对局部描 述符进行稀疏编码。利用最大池化操作聚合生成了特 征向量,根据 Minkowski 距离来衡量查询图像与数据 库图像的相似性,提升了检索准确率。文献[5]中通过 密度聚类技术进行 SIFT 特征聚类,并构造视觉词典, 再通过卡方模型分析单词与对象元素的相似性,消除 无对象信息的单词,最后基于图结构对检索结果进行 排序。该方案易受到噪声的干扰,对色彩相似图像的 误检率较高。文献[6]算法利用视觉词典构建视觉单词 直方图与倒排序索引文件,并引入 Euclidean 距离计 算相似性。该算法提取的 ROOTSIFT 无视觉不变特 性。文献[7]中通过特征值的能量布局完成对子空间的 转换,并引入 Fisher 规则,得到最优投射的哈希方 程。通过引入 RANSAC 对初始返回的图像进行优化, 提高了检索性能。由于其没有很好图像像素点的空间 信息,对全局特征利用不足,因此对于相似图像易出 现错误。



a 查询图像







b 文献[5]算法







c 文献[6]算法







图 3 不同算法的检索结果 Fig.3 Retrieval results of different algorithms

Corel 图像集中通过不同算法得到的 P-R 曲线见 图 4。P-R 曲线可综合反映算法的检索性能,从图 4 中看出, 文中算法的 P-R 曲线表现最优, 稳定性最 好。对 Corel 中 10 种类型图像的检索精度进行测试, 见图 5, 可知, 文中算法的检索精度高于其他 3 种对 照组。

为了进一步表示算法的性能,在 Corel 中,根据 返回图像数量 K(K=20, K=40)的不同对算法的效 率进行测试,结果见表 1。在 K=20 时,所提算法时 耗为 4.65 s, 而文献[5]、文献[6]、文献[7]的时耗分 别为 9.28, 11.98, 6.52 s; 在 K=40 时, 文中算法的 时耗为 9.88 s, 而文献[5]、文献[6]、文献[7]的时耗 分别为 30.64, 51.13, 18.24 s, 由此可见在相同条件 下, 文中提出的算法耗时最低。文中算法采用了基于  $\ell_0$ 稀疏约束的 NMF 对视觉词典进行增强,因此加快



图 5 不同类别的精度计算 Fig.5 Accuracy calculation of different classes

算法消耗时间对比 表 1 Tab.1 Algorithm consumption time comparison

算法	时间/s	
	<i>K</i> =20	<i>K</i> =40
文献[5]算法	9.28	30.64
文献[6]算法	11.98	51.13
文献[7]算法	6.52	18.24
文中算法	4.65	9.88

了收敛速度,减少了收敛误差。文献[5]、文献[6]中 涉及了高维数据的近邻查找问题,因此效率较低。文 献[7]中,由于在 RANSAC 中需要额外消耗时间,故 时间稍长。

### 5 结语

文中提出了一种无监督的视觉词典学习方案,在 对局部图像进行编码时大大减少了近似误差。通过结 合 NMF 和稀疏编码框架,学习了一个精确的视觉词 典。在编码阶段,提出了 ℓ<sub>0</sub>-约束 NMF 框架形成的视 觉词典,对图像集的局部描述符进行稀疏编码,通过局部 视觉词典对图像局部描述符进行稀疏编码,通过局部 聚合描述符向量(VLAD)与 Fisher 向量,利用最大 池化操作聚合生成特征值。再通过 Minkowski 距离来 计算特征值的相似性。实验结果表明,与现有的图像 检索系统相比,避免了检索算法中的语义鸿沟问题, 改进后的算法大大提高了检索性能。

通过增强的视觉词典对图像进行编码,充分考虑 了特征分布的统计信息,从而提高了检索性能。实验 数据表明,文中所提算法收敛速度快于其他算法,同 时整体性能也较高。

#### 参考文献:

- LI Li, LIN Feng, JUN Wu. Exploiting Global and Local Features for Image Retrieval[J]. Journal of Central South University, 2018, 25(2): 259–276.
- [2] 汪慧兰, 毛晓辉, 杨晶晶. 融合小波变换和 SIFT 特征的商标检索方法[J]. 计算机技术与发展, 2015, 21(4): 89—92.
  WANG Hui-lan, MAO Xiao-hui, YANG Jing-jing. Integration of Wavelet Transform and SIFT Features of Trademark Retrieval Method[J]. Computer Technology and Development, 2015, 37(7): 1654—1659.
- [3] HENG Fan, XIANG Jin-hai. Robust Visual Tracking With Multitask Joint Dictionary Learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27(5): 1018–1030.
- [4] 常振超,陈鸿昶,黄瑞阳.基于非负矩阵分解的半监督动态社团检测[J].通信学报,2016,37(2):131—142.
   CHANG Zhen-chao, CHEN Hong-chang, HUANG

Rui-yang. Semi-supervised Dynamic Community Detection Based on Non-negative Matrix Factorization[J]. Journal on Communic ations, 2016, 37(2): 131–142.

[5] 柯圣财,李弼程,陈刚.一种基于视觉词典优化和查询扩展的图像检索方法[J].自动化学报,2018,44(1): 99—105.
KE Sheng-Cai, LI Bi-Cheng, CHEN Gang. Image Retrieval with Enhanced Visual Dictionary and Query Expansion[J]. Acta Automatica Sinica, 2018, 44(1): 99–105.

- [6] LIU Bao-di, GUI Liang-ke, WANG Yu-ting. Class Specific Centralized Dictionary Learning for Face Recognition[J]. Multimedia Tools and Applications, 2017, 76(3): 4159–4177.
- [7] ZHU Song-hao, JIN Dong-liang, LIANG Zhi-wei. Integration of Semantic and Visual Hashing for Image Retrieval[J]. Journal of Visual Communication & Image Representation, 2017, 44(12): 229–235.
- [8] 姜磊, 王彤. 基于杂波数据 Frobenius 范数拟合的阵元误差估计方法[J]. 系统工程与电子技术, 2015, 37(12): 2713—2718.
  JIANG Lei, WANG Tong. Array Error Estimation Based on Frobenius Norm Fitting of Clutter Data[J]. Systems Engineering and Electronics, 2015, 37(12): 2713—2718.
- [9] RICCARDO Artoni, PATRICK Richard. Torsional Shear Flow of Granular Materials: Shear Localization and Minimum Energy Principle[J]. Computational Particle Mechanics, 2018, 5(1): 3–12.
- [10] LEE Seokjin, PANG Hee-Suk. Multichannel Non-negative Matrix Factorisation Based on Alternating Least Squares for Audio Source Separation System[J]. Electronics Letters, 2015, 51(3): 197–198.
- [11] 赵仲秋,季海峰,高隽.基于稀疏编码多尺度空间潜 在语义分析的图像分类[J]. 计算机学报,2014,37(6): 1251—1260.
  ZHAO Zhong-qiu, JI Hai-feng, GAO Jun. Sparse Coding Based Multi-scale Spatial Latent Semantic Analysis for Image Classification[J]. Chinese Journal of Computers, 2014, 37(6): 1251—1260.
- [12] LONG Hai-xia, ZHUO Li, QU Pan-ling. Weak Correlation Dictionary Construction Method for Sparse Coding[J]. Journal of Shanghai Jiaotong University (Science), 2017, 22(1): 77–81.
- [13] CHEN Bo-lun, LI Fen-fen, CHEN Sen-bo. Link Prediction Based on Non-negative Matrix Factorization[J]. Plos One, 2017, 12(8): 1—18.
- [14] 刘晓晶,印兴耀,吴国忱. 基于正交匹配追踪算法的 叠前地震反演方法[J]. 石油地球物理勘探, 2015, 50(5): 925—935.
  LIU Xiao-jing, YIN Xing-yao, WU Guo-chen. Prestack Seismic Inversion Based on Orthogonal Matching Pursuit Algorithm[J]. OGP, 2015, 50(5): 925—935.
- [15] ABD Mohamed, El Aziz, WAEL Khidr. Nonnegative Matrix Factorization Based on Projected Hybrid Conjugate Gradient Algorithm[J]. Signal, Image and Video Processing, 2015, 9(8): 1825–831.
- [16] YE Hua, MOU Qian-ying, LIU Yu-tian. Enabling Highly Efficient Spectral Discrimination-based Eigen-analysis Methods by Kronecker Product[J]. IEEE Transactions on Power Systems, 2017, 32(5): 4148—4150.

- [17] 郑秀云, 史加荣. Armijo 型线搜索下的全局收敛共轭梯 度法[J]. 山东大学学报(理学版), 2017, 52(1): 98—101.
  ZHENG Xiu-yun, SHI Jia-rong. A Globally Convergent Conjugate Gradient Method with Armijo Line Search [J]. Journal of Shandong University (Natural Science), 2017, 52(1): 98—101.
- [18] PALANCZ B, AWANE J. A Novel RANSAC Approach to Robustly Solve the 3D Similarity Transformation Problem[J]. Australian Journal of Earth Sciences, 2017, 64(4): 565–576.
- [19] CAITLYN M-parmelee. Model-based Prediction of Maximum Pool Size in the Ribbon Synapse[J]. BMC Neuroscience, 2015, 16(1): 12-21.
- [20] 顾彦慧,王道胜,王永根.基于空间短文本对象的检 索策略[J].北京大学学报(自然科学版),2016,52(1): 120—126.

GU Yan-hui, WANG Dao-sheng, WANG Yong-gen. Similar Spatial Textual Objects Retrieval Strategy[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2016, 52(1): 120—126.

- [21] 张玉兵, 宋威. 基于局部对角色差描述符的图像检索方法[J]. 计算机工程, 2016, 42(12): 248—253.
   ZHANG Yu-bing, SONG Wei. Image Retrieval Method Based on Local Diagonal Color Difference Descriptor[J].
   Computer Engineering, 2016, 42(12): 248—253.
- [22] 黄崑, 王珊珊, 王凯飞. 基于任务的图像检索相关性 评价研究[J]. 情报学报, 2016, 35(2): 115—126.
  HUANG Kun, WANG Shan-shan, WANG Kai-fei. Research on Relevance Evaluation in Task-based Image Retrieval[J]. Journal of the China Society for Scientific and Technical Information, 2016, 35(2): 115—126.