



结合 VLAD 特征和稀疏表示的图像检索

颜文, 金炜, 符冉迪

(宁波大学信息科学与工程学院, 浙江 宁波 315211)

摘要: 为了实现快速准确的图像检索目标, 提出一种结合 VLAD(局部聚合描述符)特征和稀疏表示的图像检索方法。首先, 根据图像具有结构细节丰富、局部视觉特征差异明显的特点, 提取图像的局部旋转不变 SURF 特征, 并采用局部聚合描述符方法, 构造具有旋转不变性的图像 VLAD 特征, 然后将 VLAD 特征与稀疏表示相结合, 设计基于稀疏表示的相似性检索度量准则, 实现图像的查询检索。实验结果表明, 提出方法在查准率 (precision) 及平均归一化修正检索排序等指标上, 均优于其他几种典型方法, 并具有较高的计算效率。

关键词: 图像检索; 稀疏表示; 局部聚合描述符

中图分类号: TP391

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2016308

Image retrieval based on the feature of VLAD and sparse representation

YAN Wen, JIN Wei, FU Randi

Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China

Abstract: In order to achieve the goal of fast and accurate image retrieval, an image retrieval method combining VLAD (vector of locally aggregated descriptor) feature and sparse representation was proposed. Firstly, according to the characteristics of rich structure details and obvious differences for local visual features in image, the local rotation invariant SURF feature of the image was extracted, and the local VLAD feature of the image with rotation invariance was constructed by the local aggregation descriptor method. Then, the VLAD feature was combined with the sparse representation (SR) to design the similarity retrieval metric based on SR, thus the retrieval of the image could be realized. The experimental results show that, proposed method outperforms the compared methods in terms of precision, average normalize modified retrieval rank (ANMRR) and other indicators, and it also has higher computational efficiency.

Key words: image retrieval, sparse representation, vector of locally aggregated descriptor

1 引言

近年来, 智能手机和平板电脑等移动智能终端发展迅

猛, 尤其是在多媒体处理上的能力得到了极大的提升, 利用移动端进行图像检索的需求越来越大, 比如社交网络上的以图寻图、移动购物中的商品搜索等, 如何从海量的图

收稿日期: 2016-11-07; 修回日期: 2016-12-10

基金项目: 国家自然科学基金资助项目(No.61271399); 浙江省自然科学基金资助项目 (No.LY16F010001); 宁波市自然科学基金资助项目 (No.2016A610091)

Foundation Items: The National Natural Science Foundation of China(No.61271399), The Natural Science Foundation of Zhejiang Province of China(No.LY16F010001), The Natural Science Foundation of Ningbo of China(No.2016A610091)

像库中有效地检索到目标图像成为亟待解决的难题。基于内容的图像检索(content-based image retrieval, CBIR)技术作为解决该难题的有效手段^[1],有望在移动终端的图像检索中得到广泛应用。

基于内容的移动图像检索系统的性能取决于图像特征的有效表示和与之匹配的图像相似性度量方法。传统的方法通过提取图像的全局视觉特征(如颜色、纹理、形状等)^[2,3]来进行检索,虽然具有提取复杂度低、顽健性强的优点,但当图像局部细节丰富、结构差异明显时,这种基于全局计算提取的纹理和颜色特征难于很好地表示图像的局部视觉差异。局部聚合描述符(vector of locally aggregated descriptor, VLAD)^[4]特征基于视觉词袋模型的理念,通过聚合局部旋转不变特征(scale invariant feature transform, SIFT)^[5]形成,它对局部细节丰富、结构差异明显的图像具有很强的描述能力。然而 SIFT 特征的提取需要很高的计算性能,这与移动设备有限的计算性能相矛盾。SURF(speeded up robust feature)^[6]特征与 SIFT 特征相比在可重复性、区分性、顽健性等方面具有优势,且特征维度仅有 SIFT 特征的一半,同时计算效率明显比 SIFT 高,因此本文基于 SURF 特征形成一种改进的 VLAD 特征,运用到移动图像检索中。

同时,有效的图像相似性度量方法是实现移动图像检索的另一关键环节。得益于稀疏表示在信号处理领域的诸多成功应用,特别是稀疏表示分类(sparse representation classification, SRC)识别算法优异的模式识别能力^[7],本文提出一种改进稀疏表示分类的图像稀疏表示相似性度量方法。该方法使用图像库中每类图像与对应稀疏分解系数重构出检索图像,并计算类重构残差。将类重构残差倒数的绝对值作为类权重系数与稀疏系数相乘得到加权稀疏系数,加权稀疏系数向量中的系数值即代表查询图像与图像库中每一幅图像的相似性度量值。实验结果表明,结合改进 VLAD 特征和稀疏表示相似性度量的图像检索方法,能获得比传统检索方法更好的检索效果。

2 VLAD 特征

VLAD 图像特征基于图像局部视觉词袋模型的理念构造,其形成过程主要分为 3 个步骤。首先对图像库中的每幅图像进行局部旋转不变特征提取,本文提取局部旋转不变性的 SURF 特征, SURF 特征为 64 维。然后将所有图像提取的特征汇聚在一起形成一个局部视觉特征池,并使用

k -means 聚类算法对局部视觉特征池进行聚类,得到 k 个聚类中心,将聚类中心 $c_j, j \in 1, \dots, k$ 称作码字,其集合作码书。最后将一幅图像的局部旋转不变特征集 $d=\{d_1, d_2, \dots, d_n\}$ 在码书上量化得到该图像特征。量化方式为将特征集 $d=\{d_1, d_2, \dots, d_n\}$ 中的每个局部特征 d_i 与码书中最近码字 c_j 的差值 $(d_i - c_j)$ 分配给该码字 c_j , 分配完之后累加每个码字的差值 $(d_i - c_j)$ 得到向量 $x_{VLAD}[j] = \sum_{d_i \in L_{c_j}} d_i - c_j$, 其中 L_{c_j} 表示分配到码字 c_j 的局部特征集。最终将 k 个码字对应的 $x_{VLAD}[j]$ 向量连接起来, 形成一个 $D=k \times l$ 维的 VLAD 图像特征, 其中 k 表示码书中码字的数目(即聚类中心个数), l 表示局部特征 d_i 的维数。

3 基于稀疏表示的图像检索

3.1 稀疏表示理论与稀疏分类算法

随着压缩感知概念的提出和压缩感知理论的不不断发展,信号的稀疏表示(sparse representation, SR)理论逐渐成为研究热点之一,并在信号处理、图像处理、计算机视觉、模式识别领域得到广泛应用^[8]。信号的稀疏表示是指信号可以由其所所在空间中的少数基向量线性表示,通过少数的原子揭示信号的主要特征和内在结构。从数学的角度看,是将信号在某一空间中进行线性分解。若存在一个字典 $D \in R^{n \times m} (n < m)$, 给定一个向量 $y \in R^n$ 能用字典 D 的原子线性表示,其中大部分原子系数为零或者很小,那么可以说 y 能在字典 D 上实现稀疏表示,其数学表示模型为:

$$\hat{\alpha} = \arg \min \|\alpha\|_0 \quad \text{s.t.} \quad y = D\alpha \quad (1)$$

其中, α 表示分解系数, $\|\alpha\|_0$ 表示分解系数中非零系数的数目。

假定有 k 组不同类别图像, $A_i = [A_{i1}, A_{i2}, \dots, A_{in_i}]$ 是由第 i 类图像提取的特征向量组成的训练子集 (A_{in_i} 表示每幅图像的特征向量, n_i 表示第 i 类图像的数目)。那么,来自第 i 类的待测图像 y_0 能通过训练子集 A_i 内的向量线性表示为:

$$y_0 = A_i x_i \quad (2)$$

在实际分类过程中待测图像 y_0 的类别是未知的,将 k 组训练子集 A_i 构成一个全类训练特征集 $A = [A_1, A_2, \dots, A_k]$ 。那么待测图像 y_0 就可以表示为所有训练向量的线性组合。

$$y_0 = Ax \quad (3)$$

按照稀疏表示理论,组合系数 x 仅与 y_0 相对应类别的训练子集相关,因此求解组合系数 x 就转化为稀疏表示问题。理想情况下,求解的稀疏零范数解为 $x = [0, \dots, 0, x_{i1}, x_{i2},$



$\cdots, x_{i,i_1}, 0, \cdots, 0]^T, x_{i,i_2}$ 表示与待测图像 y_0 属于同一类的表示系数。然而实际情况下,由于受到噪声的干扰和稀疏分解理论的限制,求解的解在其他类别系数上也会存在较多非 0 的系数。最终分类器设计为:通过计算测试图像与每一类训练图像和对应稀疏系数的稀疏重构残差的最小值来确定测试图像的所属类别,如式(4)所示:

$$r_i(y) = \min_i \|y - y\delta_i(x)\|_2 \quad (4)$$

其中, $\delta_i(\cdot)$ 表示提取稀疏表示系数 x 中与第 i 类的所有训练图片对应的系数,而其余的系数均赋为 0,使式(4)取得最小差值 $r_i(y)$ 所属的类别即最终的识别结果。由于图像检索的目标就是在图像库中查找与待检索图像相似的图像,并按相似性程度排序输出,因此可以在实现图像分类识别的基础上通过定义相似性度量准则实现,本文基于稀疏表示分类方法,开展了图像检索的研究。

3.2 基于稀疏表示的相似性检索方法

本文利用稀疏表示进行图像的相似性计算。依据稀疏分类算法原理,首先求解检索图像在检索图像库上的稀疏表示系数,并使用检索库中每类图像与对应稀疏系数来重构检索图像,计算出每类重构图像与检索图像的重构残差,重构残差的倒数 $R_i(i \in k)$ 表示检索图像与每一类被检索图像的类型相似性,其值越大表明检索图像与这类图像越相似。同时每一类图像对应的稀疏系数值表示这一类图像里的图像与检索图像的类型相似性,系数值越大与检索图像越相似。综合考虑检索图像与被检索图像库中图像的类型相似性和类内相似性,当被检索图像与检索图像之间

类间相似性和类内相似性乘积最大时,该图像就是与检索图像最相似的图像。具体的计算方法为,将类间相似性 R_i 作为权重与对应类内的稀疏系数相乘得到一个加权稀疏系数 α_{weight}

$$\alpha_{\text{weight}} = \left[R_1 \times \underbrace{|x_1|, |x_2|, \dots}_{\hat{\partial}_1}, R_2 \times \underbrace{|x_j|, \dots}_{\hat{\partial}_2}, \dots, R_k \times \underbrace{|\dots|, |x_n|}_{\hat{\partial}_k} \right] \quad (5)$$

α_{weight} 向量中每个元素值表示检索图像 q 与被检索图像库中每一幅图像的相似性度量值。

3.3 检索过程

本文结合 VLAD 特征与稀疏表示相似性度量进行图像检索,其检索流程如图 1 所示。具体步骤如下。

步骤 1 首先提取检索图像库 M 中每幅图像的 VLAD 特征 A_i ,然后将每幅图像的 VLAD 特征汇聚在一起构成被检索图像特征字典 $A = [A_1, A_2, \dots, A_n] \in R^{m \times n}$, m 表示特征的维数, n 表示检索图像库中图像的个数。

步骤 2 提取查询图像 q 的 VLAD 特征 $A_q \in R^m$, 并利用稀疏分解算法求解 A_q 在被检索图像特征字典 A 上的稀疏表示系数 $x = [x_1, x_2, x_3, \dots, x_n]$ 。

步骤 3 将字典 A 和稀疏系数 x 划分为 k 类。

$$A = \left[\underbrace{A_1, A_2, \dots}_{D_1}, \underbrace{A_j, \dots}_{D_2}, \dots, \underbrace{A_n}_{D_k} \right] \quad (6)$$

$$x = \left[\underbrace{x_1, x_2, \dots}_{\hat{\partial}_1}, \underbrace{x_j, \dots}_{\hat{\partial}_2}, \dots, \underbrace{x_n}_{\hat{\partial}_k} \right] \quad (7)$$

步骤 4 计算每一类的重构残差 $r_i = \|A_q - D_i \times \hat{\partial}_i\|_2$, 将重

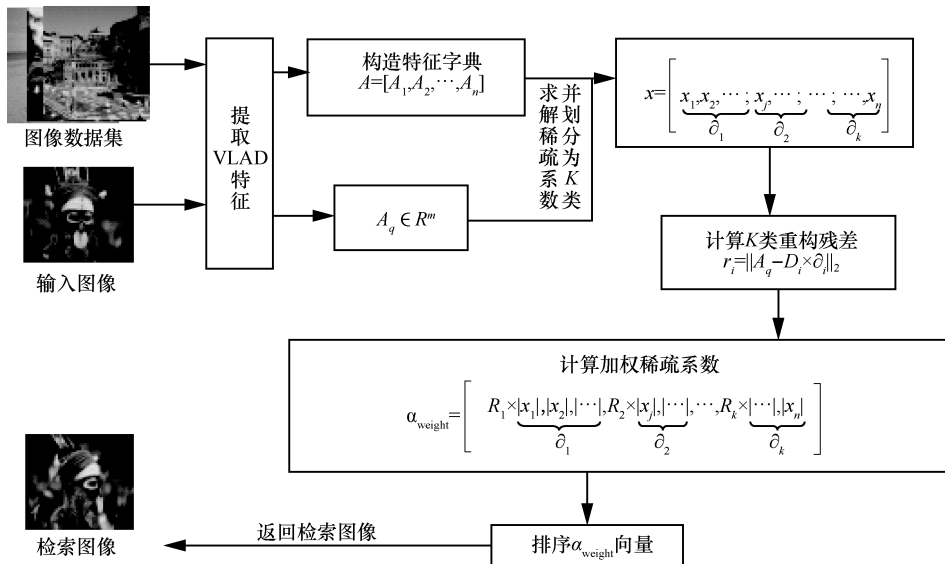


图 1 本文图像检索流程

构造残差 r_i 的倒数 R_i 作为权重,并与第 i 类稀疏系数 d_i 中每一个系数的绝对值相乘,得到一个加权稀疏系数 α_{weight} 向量。最终 α_{weight} 向量中的值代表查询图像 q 与图像库中每一幅图像的相似性度量值。

$$\alpha_{\text{weight}} = \left[R_1 \times \underbrace{|x_1|, |x_2|, \dots}_{c_1}, R_2 \times \underbrace{|x_j|, \dots}_{c_2}, \dots, R_k \times \underbrace{|\dots|, |x_n|}_{c_k} \right] \quad (8)$$

步骤 5 排序 α_{weight} 向量,返回最大相似性度量值对应的图像作为检索图像 q 最相关的图像。

4 实验结果与分析

4.1 实验图像数据集和评价指标

本文使用 Corel-1000 自然图像集作为检索图像集^[9]。该图像库包含非洲风光、海滩、建筑等 10 类图像,每类由 100 幅彩色图像构成。

为了评价图像检索方法的有效性,采用平均查准率和查准—查全率 (precision-recall) 曲线作为性能评价标准。同时,为了评价检索结果的排序情况,实验中引入了平均归一化修正检索排序 (average normalize modified retrieval rank, ANMRR)^[10] 这一 MPEG-7 标准化评价准则。ANMRR 取值在 [0,1] 之间,如果 ANMRR 取值为 0 时,说明图像库中所有相关图像全部被检索出来,反之当 ANMRR 取值为 1 时,情况刚好相反,因此 ANMRR 取值越小,检索方法的性能越好。实验中,每一类随机选取 50 幅图像作为待检索图像,10 类共进行 500 次查询,取检索结果的平均值对检索方案进行验证。

4.2 对比实验的设计及结果分析

为了检验不同检索方法的性能,将本文方法与两种典型的检索算法进行比较,方法 1 为基于局部和全局特征融合的图像检索^[11],方法 2 为基于 NSCT 及熵的旋转不变彩色图像检索算法^[12]。由于形成 VLAD 特征的视觉码本大小对检索性能有直接影响,为了选取合适的码本大小,分别将码本设置为 64、128、512 进行实验测试;并采用 GPSR (gradient projection sparse reconstruction)^[13] 进行稀疏系数的求解。

表 1 为 3 种方法的实验结果,其中计算查准率时,每次检索均采用返回 30 幅图像作为查询结果。可以看出,方法 1 在查准率及 ANMRR 指标中均表现最差,究其原因,主要是由于自然图像具有丰富的细节结构,且不同类型图像的局部视觉差异明显,而方法 1 提取的特征不能很好地

表现局部视觉差异性,同时人工简单设置不同特征加权重没有考虑不同特征的结合特性。方法 2 提取了能表示不同类型图像细节结构的旋转不变纹理特征,同时结合颜色和形状特征,获得了比方法 2 好的检索效果。本文方法的查准率及 ANMRR 指标在 3 种方法中都表现得最好,这不仅体现了基于局部 SURF 的 VLAD 特征对自然图像不同局部视觉差异性极强的区分描述能力,而且表明基于特征稀疏表示的相似性检索方法可有效提高检索精度。同时从实验结果可以看出,当使用不同码本的 VLAD 特征时,大码本检索效果比小码本差一些,可能原因是随着码本大小增加,会造成高维特征信息的冗余,使检索性能的下降,因此选取形成 VLAD 特征的码本大小为 64。为了从不同角度说明本文方法的检索性能,将码本大小设置为 64 来进行对比试验,并绘制了对应的查准—查全率曲线,如图 2 所示,可以看出方法 1 与方法 2 性能比较接近,都取得了不错的检索效果,但相对而言,本文方法取得了最好的检索效果,这也与前面的分析相吻合。

表 1 两种方法的平均查准率和 ANMRR 对比

对比方法	VLAD+SR			方法 1	方法 2
码本大小	64	128	512		
查准率	68.24	67.11	65.32	60.51	62.43
ANMRR	0.476 1	0.499 5	0.513 4	0.546 1	0.530 1

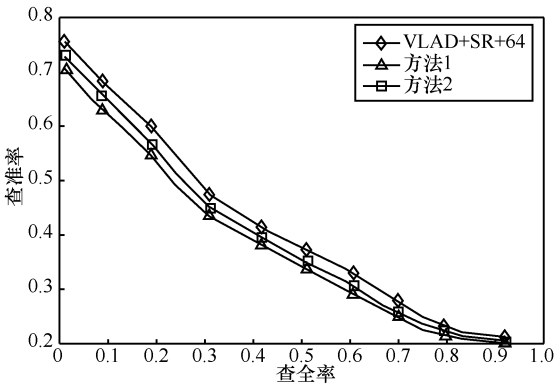


图 2 3 种方法的查准—查全率比较

移动图像检索对检索时间要求较高,因此在评价图像检索方法的整体性能时,需要综合考虑检索性能与检索时间。从上文分析可知,码本大小为 64 维时,获得了最好的检索结果,但采用 64 维码本形成的 VLAD 特征具有较高的维度 (VLAD 特征维数为 4 096 维),使得检索时间过长。特征降维虽然能减少检索时间,但是也会造成特征信息的丢失,从而使特征内在区分描述能力不足,带来了



表 2 降维之后不同特征维度检索性能比较

检索性能	4 096	2 048	1 024	512	256	128
检索耗时/s	2.21	1.56	0.91	0.79	0.70	0.67
ANMRR	0.476 1	0.480 2	0.498 7	0.504 3	0.532 4	0.551 3

检索性能的下降。因此在特征降维的过程中需要在保持特征内在区分能力的同时尽量降低特征维数,以缩短检索时间,从而取得最佳的性能平衡。为达到此目标,本文采用 PCA 降维的方法进行了实验,确定合适的降维维度,实验结果见表 2。

从表 2 中可以看到,当采用原始特征时,虽然检索准确率最好,但检索时间最长,限制了其在移动设备上的应用,而将原始特征降至 128 维时,虽然检索用时较短,但带来了检索性能的急剧降低,本文综合考虑检索性能与检索时间的平衡,认为将原始特征降至 512 维较为合适,表 3 给出了特征维数为 512 时,本文方法与两种对比方法的实验结果。

表 3 特征降维后与两种方法的比较

检索性能	VLAD+SR	方法 1	方法 2
检索耗时/s	0.79	0.65	0.74
ANMRR	0.504 3	0.546 1	0.530 1

从实验结果可以看到,经过特征降维后,本文方法的检索性能明显优于两种对比方法,而在检索效率上,本文方法的检索耗时与对比方法基本相当,表明本文方法具有一定的应用价值。

5 结束语

基于内容图像检索系统的检索效果取决于有效的图像特征表示和与之相匹配的相似度衡量方法。在特征表示方面,不同类型的图像均可能具有结构细节信息丰富,局部视觉特征差异明显的特点,SURF 具有较强的图像特征描述能力,且通过局部聚合描述符描述方法形成 VLAD 特征能更好地描述图像信息,本文基于 VLAD 特征和稀疏表示理论,提出一种图像相似性度量方法,实现图像检索。实验结果表明,相比几种较新的检索方法,本文方法在保证检索效率的同时能显著提高检索性能,有一定的应用前景。

参考文献:

[1] LIU G H, YANG J Y, LI Z Y. Content-based image retrieval using computational visual attention model [J]. Pattern

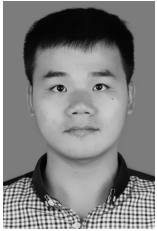
Recognition, 2015, 48(8): 2554-2566.

- [2] 孙君顶,崔江涛,毋小省,等. 基于颜色和形状特征的彩色图像检索方法[J]. 中国图象图形学报, 2004, 9(7): 820-827.
SUN J D, CUI J T, WU X S, et al. Color image retrieval based on color and shape features [J]. Journal of Image and Graphics, 2004, 9(7): 820-827.
- [3] 杨舒,王玉德. 基于 Contourlet 变换和 Hu 不变矩的图像检索算法[J]. 红外与激光工程, 2014, 43(1): 306-310.
YANG S, WANG Y D. Image retrieval algorithm based on Contourlet transform and Hu invariant moments[J]. Infrared and Laser Engineering, 2014, 43(1): 306-310.
- [4] JEGOU H, DOUZE M, SCHMID C. Aggregating local descriptors into a compact image representation [C]//IEEE Conference on Computer Vision & Pattern Recognition, Jun 23-28, 2013, Portland, Oregon, USA. New Jersey: IEEE Press, 2010: 3304-3311.
- [5] LOWE D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(60): 91-110.
- [6] BAY H, ESS A, TUYTELAARS T, et al. Speeded-up robust features(SURF) [J]. Computer Vision & Image Understanding, 2008, 110(3): 346-359.
- [7] WRIGHT J, YANG A Y, GANESH A, et al. Robust face recognition via sparse representation [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2009, 31(2): 210-227.
- [8] ZHANG Z, XU Y, YANG J, et al. A survey of sparse representation: algorithms and applications [J]. IEEE Access, 2015(3): 1.
- [9] LI J, WANG J Z. Automatic linguistic indexing of pictures by a statistical modeling approach [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2003, 25(9): 1075-1088.
- [10] MANJUNATH B S, SALEMBIER P, SIKORA T. Introduction to MPEG7: multimedia content description interface [M]. New York: Wiley, 2002.
- [11] 汪华章,何小海,宰文姣. 基于局部和全局特征融合的图像检索[J]. 光学精密工程, 2008, 16(6): 1098-1104.
WANG H Z, HE X H, ZAI W J. Image retrieval based on combining local and global features [J]. Optics and Precision Engineering, 2008, 16(6): 1098-1104.
- [12] 赵晓丽,王国中. 基于 NSCT 及熵的旋转不变彩色图像检索算法[J]. 光电子·激光, 2014(1): 186-191.
ZHAO X L, WANG G Z. Rotation-invariant color image

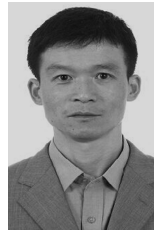
retrieval algorithm based on NSCT and entropy [J]. Journal of Optoelectronics-Laser, 2014(1): 186-191.

- [13] FIGUEIREDO M A T, NOWAK R D, WRIGHT S J. Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems[J]. IEEE Journal of Selected Topics in Signal Processing, 2008, 1(4): 586-597.

[作者简介]



颜文(1990-),男,宁波大学信息科学与工程学院硕士生,主要研究方向为图像处理、模式识别。



金炜(1969-),男,博士,宁波大学信息科学与工程学院副教授,主要研究方向为数字图像处理。



符冉迪(1971-),男,宁波大学信息科学与工程学院副教授,主要从事数字图像处理、模式识别方面的研究工作。