

## 类别约束下的低秩优化特征字典构造方法

吕 煊<sup>1</sup>, 刘玉淑<sup>2\*</sup>, 丁洪富<sup>1</sup>, 李爱迪<sup>1</sup>

(1. 重庆市国土资源和房屋勘测规划院, 重庆 400020; 2. 齐鲁工业大学 电气与自动化学院, 济南 250353)

(\* 通信作者电子邮箱 654080757@qq.com)

**摘 要:**字典模型(BOW)是一种经典的图像描述方法,模型中特征字典的构造方法至关重要。针对特征字典构造问题,提出了一种类别约束下的低秩优化特征字典构造方法 LRC-DT,通过低秩优化的方法使训练出来的特征字典在描述同类图像时表示系数矩阵的秩相对较低,从而将类别信息引入到字典学习中,提高字典对图像描述的可分辨性。在标准公测库 Caltech-101 和 Caltech-256 上的实验结果表明:将 SPM、稀疏编码下的 SPM(ScSPM)、局部线性编码(LLC)和线性核函数的 SPM(LSPM)编码方法中的特征字典替换为加入低秩约束(LRC)的特征字典后,随着训练样本数目增多,字典模型分类准确率与未引入低秩约束的方法相比有所提高。

**关键词:**字典模型;低秩优化;低秩描述;图像描述;图像分类

**中图分类号:** TP391.41 **文献标志码:** A

### Low-rank optimization characteristic dictionary training approach with category constraint

LYU Xuan<sup>1</sup>, LIU Yushu<sup>2\*</sup>, DING Hongfu<sup>1</sup>, LI Aidi<sup>1</sup>

(1. Chongqing Land Resources Housing Surveying and Planning Institute, Chongqing 400020, China;

2. School of Electrical Engineering and Automation, Qilu University of Technology, Jinan Shandong 250353, China)

**Abstract:** Bag Of Words (BOW) is a classical approach of image description, and the method of constructing the characteristic dictionary in this model is very important. A category constrained low-rank optimization characteristic dictionary training approach named LRC-DT was proposed for the characteristic dictionary construction. Through the low-rank optimization, the rank of the coefficient matrix constructed by same category images was minimized. Then the classification information was introduced into the characteristic dictionary learning to improve the identifiability of characteristic dictionary for image description. Some experiments were conducted on two standard image databases including Caltech-101 and Caltech-256, and the characteristic dictionary of SPM (Spatial Pyramid Matching), ScSPM (Sparse codes SPM), LLC (Locality-constrained Linear Coding) and LSPM (Linear SPM) were replaced by constrained low-rank optimization characteristic dictionary. The experimental results show that the proposed method can consistently offer better performance than not employing the category constrained low-rank optimization, its classification accuracy is improved with the increase of the training sample number.

**Key words:** Bag Of Words (BOW); low-rank optimization; low-rank representation; image representation; image classification

## 0 引言

字典模型(Bag Of Words, BOW)<sup>[1]</sup>是一种通过统计字典中特征单词出现的频次来描述一幅图像直方图式的统计特征图像描述方法。目前,有关图像特征描述的方法很多,如文献[2]中所述,对应于不同的特征检测方法、特征描述方法以及特征的组织与表达方法。字典模型因其较强的图像特征描述能力,在图像检索<sup>[3-4]</sup>、分类<sup>[5]</sup>、编码<sup>[6-7]</sup>等领域均得到了广泛的应用。字典模型在不同的应用领域有着不同的创建和使用方法。在图像去噪、超分辨率图像重构时,通常将图像划分为大小相同的小块,文献[8]通过对这些局部小块进行特征分析,构造能够描述局部小块的特征字典。在图像分类中通常对整幅图像作特征分析,一般由特征检测、特征描述、特征聚类三部分组成,由各聚类中心构成特征字典<sup>[9-12]</sup>。

在基于字典模型的图像分类问题中,如何创建一个稳定、分辨性高、描述能力强的特征字典至关重要。好的特征字典可以保证图像重构误差最小的情况下,最大化图像描述的可分辨性。构造字典模型时通常有两种思路:一种是利用所有类别的图像构造一个特征字典;另一种是对每一类图像构造单独的特征字典,从而构造多个特征字典。针对单一字典没有充分利用类别信息,而构造多类字典,当类别特别多时计算代价非常高而且容易丧失对分类有益的各类图像中的共性特征信息等问题,本文提出了一种带类别约束的利用低秩优化方式创建特征字典的方法(Low-Rank Constrained Dictionary Training, LRC-DT),通过引入低秩约束(Low-Rank Constraint, LRC)的思想,使得特征字典在表示同类图像时生成的表示系数尽可能相似,从而将类别信息引入到字典学习中,提高字典对图像描述的可分辨性。

收稿日期:2014-03-19;修回日期:2014-05-09。 基金项目:国土资源部公益性项目(201311006)。

作者简介:吕煊(1982-),男,山东淄博人,工程师,博士,主要研究方向:图像分类、数据挖掘; 刘玉淑(1982-),女,山东淄博人,讲师,博士,主要研究方向:数字图像处理、模式识别; 丁洪富(1974-),男,重庆人,正高级工程师,主要研究方向:地理信息; 李爱迪(1979-),男,四川人,高级工程师,主要研究方向:地理信息系统。

## 1 相关工作

常见的用于图像分类的字典方法有 SPM (Spatial Pyramid Matching)<sup>[11]</sup> 和 BoF (Bag-of Features)<sup>[13]</sup>, 以及对 BoF 方法的诸多扩展, 如文献[14-15]中的局部生成模型、文献[16]的潜在生成模型、文献[17-18]中的辨别字典学习等。此外, 字典模型在目标监测<sup>[19]</sup>、特征描述子构造<sup>[20]</sup>中也有广泛的应用。在这些方法中, 特征字典起到了非常关键的作用。文献[9]最早提出了“关键特征字典”这个术语, 也是最早将该方法用于图像分类的文献之一。文献中通过仿射协变区域和尺度不变特征转换 (Scale-Invariant Feature Transform, SIFT) 特征描述子的方法进行特征描述, 利用  $K$ -means 聚类得到特征字典, 然后用贝叶斯分类器和支持向量机 (Support Vector Machine, SVM) 分类器分别进行了分类实验。之后文献[10]在此之上进行了改进, 采用了规范化的非仿射变换的 SIFT 局部描述, 用 Boost 代替 SVM, 并加入了细微的几何一致性约束。文献[12]更系统地研究了特征字典模型方法, 并对特征检测方法和特征描述子, 如 SIFT、RIFT (Rotation Invariant Feature Transform), 以及 SVM 核函数进行了实验。文献[21-22]将稀疏优化应用到了字典模型方法中, 因其构造的字典描述可分辨性高, 成为了近年来主流的字典模型构造方法之一。

利用特征字典模型来进行图像描述的思想是: 首先, 利用字典中特征向量来表示图像中检测到的每个特征向量; 然后根据图像中所有特征在字典模型下对应的字典表示系数来描述一幅图像; 最后根据字典描述特征向量构造分类器。

字典模型方法中需要考虑的关键问题有:

- 1) 特征的选取, 包括特征检测方法和特征描述方法;
- 2) 特征提取后, 特征字典的构造方法, 如传统的聚类方法、稀疏约束方法、 $K$  奇异值分解 ( $K$ -means Singular Value Decomposition, KSVD)<sup>[23-24]</sup> 方法等;
- 3) 特征字典对新入特征的代表方法, 即向量量化的方法, 如最近邻编码方式、弹性加权方式以及稀疏编码方式等;
- 4) 向量量化后 Pooling 的方法, 即得到图像中所有特征点在字典模型下的表示系数后, 如何构造整个图像的特征描述。
- 5) 利用特征字典描述后的特征向量构造分类器进行识别。

本文中用到的数学符号及其含义如下:

$D \in \mathbf{R}^{n \times K}$  表示具有  $K$  个原子的特征字典矩阵, 每个原子为  $n$  维特征向量;

$X = [x_1, x_2, \dots, x_M] \in \mathbf{R}^{n \times M}$  表示图像中的特征集合, 由  $M$  个  $n$  维特征向量组成的特征矩阵, 向量  $x_i \in \mathbf{R}^{n \times 1}$  ( $i = 1, 2, \dots, m$ ) 为第  $i$  个特征的特征向量;

$U = [u_1, u_2, \dots, u_M] \in \mathbf{R}^{K \times M}$  表示  $X$  在字典  $D$  下的系数表示矩阵。

若未特别指明, 所有向量均为列向量, 取转置后表示行向量, 如  $y_i \in \mathbf{R}^{n \times 1}$ ,  $y_i^T \in \mathbf{R}^{1 \times n}$ ,  $y_i = Du_i$  表示输入信号  $y_i \in \mathbf{R}^{n \times 1}$  可由字典矩阵  $D$  中  $K$  列原子线性表达。

### 1.1 传统字典构造方法

在传统的特征字典构造方法中, 文献[11]通过对所有图像中检测到的特征进行聚类 (如  $K$ -means) 获得特征字典, 聚类中心的个数即为特征字典中单词个数。若特征集合  $X = [x_1, x_2, \dots, x_M] \in \mathbf{R}^{d \times M}$  是  $M$  个  $d$  维特征空间下的特征点,  $V = [v_1, v_2, \dots, v_K]^T \in \mathbf{R}^{K \times d}$  是  $K$  个聚类中心即大小为  $K$  的特

征字典, 则量化问题为求最解优化式(1):

$$\min_{U, V} \sum_{m=1}^M \|x_m - u_m V\|^2 \quad (1)$$

$$\text{s. t. } \text{Card}(u_m) = 1, |u_m| = 1, u_m \geq 0, \forall m$$

其中:  $U = [u_1, u_2, \dots, u_M]^T$ , 向量  $u_m$  中元素表示特征点  $x_m$  与  $K$  个聚类中心的对应关系;  $\text{Card}(u_m) = 1$  表示  $u_m$  中只有一个元素非零,  $u_m$  中元素都是非负的;  $|u_m|$  表示  $l_1$  范数。

传统方法中使用最近邻的特征单词来表示新入特征, 然后计算字典中单词的统计直方图用来表示图像。该方法可以在一定程度上反映图像中的特征分布情况, 但是仅用最近邻单词来表示新入特征, 重构误差较大, 对新入特征的描述不够准确。后来, 基于弹性权值 (Soft-Weighting) 的方式对传统方法进行了改进, 它利用字典中与输入信号特征向量最近的  $k$  个原子来表示输入信号,  $k$  个原子的权值由原子与信号间的相似度来决定, 相似度越大则权值越大。该方法虽然简单、易实现, 与  $K$ -means 方法相比对信号的重构误差更小, 但该描述方式的可分辨性较差。

### 1.2 稀疏编码特征字典

与  $K$ -means 编码方式和基于弹性权值的方法不同, 稀疏编码方法中新特征的描述由稀疏个特征单词来表示。文献[21]提出了一种稀疏编码方式训练特征字典的方法——Sparse Codes 下的 SPM 图像描述方法 (以下简称 ScSPM), 该方法通过放宽对  $u_m$  的  $\text{Card}(u_m) = 1$  约束, 改为  $l_0$  范数约束, 利用稀疏个原子来线性表示信号, 使得字典  $D$  对  $X$  的描述更加精准, 求解问题变为式(2)的稀疏编码方式:

$$\min_{U, V} \sum_{m=1}^M \|x_m - u_m V\|^2 + \lambda \|u_m\|_{l_1} \quad (2)$$

$$\text{s. t. } \|v_k\| \leq 1, \forall k = 1, 2, \dots, K$$

$l_1$  范数定义为向量  $u_m$  中元素绝对值之和, 字典  $V$  中有  $K$  个单词;  $l_2$  范数约束项  $\|v_k\| \leq 1$  的作用是避免出现无意义的特征单词结果;  $V$  初始化为  $K$ -means 聚类求得的初始字典。式(2)可通过固定  $U$  求解  $V$ , 固定  $V$  再求解  $U$  的交替优化方式求解。稀疏编码方法通过求得的稀疏个单词来表示新的特征, 比传统方法重构误差小, 而且可分辨性也更好。

### 1.3 局部约束线性编码方式

文献[22]提出了一种局部约束线性编码 (Locality-constrained Linear Coding, LLC), 该方法是基于局部性可至稀疏性思想提出的。与稀疏编码方式相比, 该方法有解析解, 计算复杂度为  $O(K+K)$ ,  $K$  为词典中原子的个数。方法核心为将式(2)中的  $l_1$  范数约束替换为局部线性约束如下:

$$\min_C \sum_{m=1}^M \|x_m - u_m V\|^2 + \lambda \|d_m \odot c_m\|^2 \quad (3)$$

$$\text{s. t. } 1^T c_m = 1, \forall m$$

其中:  $\odot$  表示对应元素相乘,  $d_m \in \mathbf{R}^d$  为对应于  $V$  中每个基向量的局部权重, 根据输入描述子  $x_m$  与基向量的距离算得

$$d_m = \exp(\text{dist}(x_m, V)/\sigma) \quad (4)$$

其中:

$$\text{dist}(x_m, B) = [\text{dist}(x_m, b_1), \text{dist}(x_m, b_2), \dots, \text{dist}(x_m, b_K)]^T \quad (5)$$

$\text{dist}(x_m, b_j)$  为  $x_m$  与  $b_j$  间的欧氏距离,  $\sigma$  用于调节权重步长。文献[22]中给出了  $c_m$  的解析解如下:

$$\tilde{c}_m = (C_m + \lambda \text{diag}(d)) \setminus 1 \quad (6)$$

$$c_m = \tilde{c}_m / 1^T \tilde{c}_m \quad (7)$$

其中:

$$C_m = (V - 1x_m^T)(V - 1x_m^T)^T \quad (8)$$

表示数据协方差矩阵。解析解的给出使得 LLC 在实际应用中可以更加快速地计算。该方法在引入稀疏约束的同时,将稀疏表达集中在了局部邻近的相似的稀疏个单词上面,而稀疏约束方法(ScSPM)通常会使得稀疏表达分散在差别较大的稀疏个单词上面。

## 2 类别约束下的低秩优化特征字典构造

### 2.1 基本思想

特征字典的构造至关重要,好的特征字典能够在尽量减少新入特征重构误差的同时,满足同类图像的字典表示更相近,而不同类图像的字典表示相差较大。式(1)~(3)中构造特征字典时没有用到图像类别信息,而类别信息有助于构造出具有更好分辨性的字典。字典对图像的描述表现在描述系数上,即字典中各原子线性组合的系数,同一类的图像在字典模型描述下系数矩阵是相近的,而不同类图像在字典模型描述下的系数矩阵差别是比较大的。

低秩思想应用很广泛,如图像矫正<sup>[25]</sup>、描述子空间结构<sup>[26]</sup>、特征提取<sup>[27]</sup>、流行优化<sup>[28]</sup>、矩阵复原<sup>[29]</sup>、子空间分割<sup>[30]</sup>等。在图像矫正中,按照变换参数 $\tau$ 矫正后图像的秩会变小,因为矫正后图像中各列(或各行)数据相似或相同的变多,因此图像的秩变小。求解该问题可转化为求解如下优化问题:

$$\min_{I^0, E, \tau} \text{rank}(I^0) + \gamma \|E\|_0 \quad (9)$$

$$\text{s.t. } I \circ \tau = I^0 + E$$

式(9)中的求秩部分不利于该优化问题的求解,通常会近似替换为核范数求解:

$$\min_{I^0, E, \tau} \|I^0\|_* + \gamma \|E\|_0 \quad (10)$$

$$\text{s.t. } I \circ \tau = I^0 + E$$

借鉴低秩的思想,本文将低秩优化应用到字典的创建中,即通过最小化同类图像字典描述下系数矩阵的秩,引入类别信息约束,使得同类图像的字典表示更相近。为达到此目的,构造了如下最优化问题,求解低秩约束特征字典 $D$ :

$$\min \sum_{i=1}^C (\|Z_i\|_* + \lambda \|E_i\|_{2,1}) \quad (11)$$

$$\text{s.t. } X_i = DZ_i + E_i, i = 1, 2, \dots, C$$

$$\sum_{m=1}^p d_{mn}^2 \leq 1, n = 1, 2, \dots, k$$

其中: $\|Z_i\|_*$ 为矩阵的核范数,定义为该矩阵奇异值之和; $\|E\|_{2,1} = \sum_j \sqrt{\sum_i [E]_{ij}^2}$ 为 $l_2/l_1$ 范数,是规则项。共 $C$ 类 $Z_i$ 为第 $i$ 类图像在字典 $D$ 下的系数矩阵,不等式约束的作用是避免出现无意义结果。

### 2.2 低秩优化问题求解

本文采用增广型拉格朗日方法求解最优化式(11),过程如下:

令 $Z_i = J_i$ ,式(11)转换为如下增广型拉格朗日乘子式:

$$\min \sum_{i=1}^C (\|J_i\|_* + \lambda \|E_i\|_{2,1} + \text{tr}\langle W_i, Z_i - J_i \rangle + \text{tr}\langle Y_i, X_i - DZ_i - E_i \rangle + 0.5\mu \|X_i - DZ_i - E_i\|_F^2 + 0.5\mu \|Z_i - J_i\|_F^2) + \text{tr}(B(D^T D - I)) \quad (12)$$

其中 $B$ 为对角矩阵。

更新 $J_i$ :

$$\|J_i\|_* + 0.5\mu(2\text{tr}\langle W_i, Z_i - J_i \rangle/\mu + \|Z_i - J_i\|_F^2 + (\|W_i\|^2/\mu^2) - (\|W_i\|^2/\mu^2)) \Rightarrow \|J_i\|_* + 0.5\mu \|J_i - (Z_i + W_i/\mu)\|_F^2 \quad (13)$$

对 $J_i$ 的更新转化为求解优化式(14):

$$J_i^* = \arg \min \|J_i\|_* / \mu + 0.5 \|J_i - (Z_i + W_i/\mu)\|_F^2 \quad (14)$$

式(14)可由奇异值阈值划分法求得,定义奇异值阈值划分操作如下:

$$S_\varepsilon[x] = \begin{cases} x - \varepsilon, & x > \varepsilon \\ x + \varepsilon, & x < -\varepsilon \\ 0, & \text{其他} \end{cases} \quad (15)$$

其中 $x \in \mathbf{R}$ 并且 $\varepsilon > 0$ ,该方法容易扩展到向量和矩阵模式下的计算。式(14)中的优化问题可通过下式求解:

$$US_\varepsilon[S]V^T = \arg \min_x \varepsilon \|X\|_* + 0.5 \|X - W\|_F^2 \quad (16)$$

其中 $USV^T$ 为 $W$ 的奇异值分解(Singular Value Decomposition, SVD)。

更新 $Z_i$ :

$$\begin{aligned} & \text{tr}\langle W_i, Z_i - J_i \rangle + \text{tr}\langle Y_i, X_i - DZ_i - E_i \rangle + \\ & 0.5\mu \|X_i - DZ_i - E_i\|_F^2 + 0.5\mu \|Z_i - J_i\|_F^2 \Rightarrow \\ & 0.5\mu \{ 2\text{tr}\langle W_i, Z_i - J_i \rangle/\mu + \|Z_i - J_i\|_F^2 + \\ & (\|W_i\|^2/\mu^2) - (\|W_i\|^2/\mu^2) + \\ & 2\text{tr}\langle Y_i, X_i - DZ_i - E_i \rangle/\mu + \|X_i - DZ_i - E_i\|_F^2 + \\ & (\|Y_i\|^2/\mu^2) - (\|Y_i\|^2/\mu^2) \} \Rightarrow \\ & 0.5\mu \{ \|Z_i - J_i + W/\mu\|_F^2 + \|X_i - DZ_i - E_i + Y_i/\mu\|_F^2 \} \end{aligned} \quad (17)$$

式(17)对 $Z_i$ 求偏导:

$$\delta \frac{\mu}{2} \left( \left\| Z_i - J_i + \frac{W}{\mu} \right\|_F^2 + \left\| X_i - DZ_i - E_i + \frac{Y_i}{\mu} \right\|_F^2 \right) \frac{\delta Z_i}{\delta Z_i} = 0 \quad (18)$$

得到:

$$Z_i = (1 + D^T D)^{-1} (J_i + D^T (X_i - E_i) + (D^T Y_i - W_i)/\mu) \quad (19)$$

更新 $E_i$ :

$$\begin{aligned} & \lambda \|E_i\|_{2,1} + 0.5\mu \|X_i - DZ_i - E_i + Y_i/\mu\|_F^2 + \\ & \text{tr}\langle Y_i, X_i - DZ_i - E_i \rangle \Rightarrow \\ & \lambda \|E_i\|_{2,1} + 0.5\mu \|E_i - (X_i - DZ_i + Y_i/\mu)\|_F^2 \end{aligned} \quad (20)$$

$$E_i^* = \arg \min \lambda \|E_i\|_{2,1} / \mu + 0.5 \|E_i - (X_i - DZ_i + Y_i/\mu)\|_F^2 \quad (21)$$

式(21)可根据如下引理<sup>[31]</sup>求解:

给定矩阵 $Q = [q_1, q_2, \dots, q_i, \dots]$ ,若 $W^*$ 为下式的最优解:

$$\min_W \lambda \|W\|_{2,1} + 0.5 \|W - Q\|_F^2 \quad (22)$$

则 $W^*$ 的列向量可由下式求得:

$$[W^*]_{:,i} = \begin{cases} \frac{\|q_i\|_2 - \lambda}{\|q_i\|_2} q_i, & \lambda < \|q_i\|_2, \\ 0, & \text{其他} \end{cases} \quad (23)$$

更新 $D$ :

$$\begin{aligned} & \sum_{i=1}^C (\text{tr}\langle Y_i, X_i - DZ_i - E_i \rangle + 0.5\mu \|X_i - DZ_i - E_i\|_F^2) + \\ & \text{tr}(B(D^T D - I)) \Rightarrow \\ & \sum_{i=1}^C 0.5\mu (\|DZ_i - X_i + E_i - Y_i/\mu\|_F^2) + \end{aligned}$$



$$\text{tr}(\mathbf{B}(\mathbf{D}^T \mathbf{D} - \mathbf{I})) \quad (24)$$

式(24)对 $\mathbf{D}$ 求偏导得:

$$\frac{\delta \sum_{i=1}^c 0.5\mu(\|\mathbf{D}\mathbf{Z}_i - \mathbf{X}_i + \mathbf{E}_i - \mathbf{Y}_i/\mu\|_F^2) + \text{tr}(\mathbf{B}(\mathbf{D}^T \mathbf{D} - \mathbf{I}))}{\delta \mathbf{D}} = 0 \quad (25)$$

求得 $\mathbf{D}$ 为:

$$\mathbf{D} = \sum_{i=1}^c (\mathbf{X}_i - \mathbf{E}_i + \mathbf{Y}_i/\mu) \mathbf{Z}_i^T \left( \sum_{i=1}^c \mu \mathbf{Z}_i \mathbf{Z}_i^T + 2\mathbf{B} \right)^{-1} \quad (26)$$

更新乘子:

$$\begin{cases} \mathbf{W}_i = \mathbf{W}_i + \mu(\mathbf{Z}_i - \mathbf{J}_i) \\ \mathbf{Y}_i = \mathbf{Y}_i + \mu(\mathbf{X}_i - \mathbf{D}\mathbf{Z}_i - \mathbf{E}_i) \\ \mathbf{B} = \mathbf{B} + \mu(\mathbf{D}^T \mathbf{D} - \mathbf{I}) \end{cases} \quad (27)$$

### 2.3 基于低秩优化特征字典的图像描述

在求得特征字典 $\mathbf{D}$ 之后,固定字典求解新描述子 $\mathbf{x}_m$ 在 $\mathbf{D}$ 下的描述,即固定 $\mathbf{D}$ 的前提下求解式(28):

$$\min_{\mathbf{u}_m} \|\mathbf{x}_m - \mathbf{u}_m \mathbf{D}\|^2 + \lambda \|\mathbf{u}_m\|_{l_1} \quad (28)$$

得到图像中所有描述子在字典 $\mathbf{D}$ 下的稀疏表示系数 $\mathbf{u}_m$ 后,采用文献[11]中的SPM分块方法,对第 $l$ 层中的第 $k$ 个子块中特征描述子集合采用最大化池的方法构造 $\mathbf{X}$ 的特征向量 $\mathbf{F}_l = [\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_k]$ ,其中,

$$\mathbf{f}_j = \max\{|\mathbf{u}_{1j}|, |\mathbf{u}_{2j}|, \dots, |\mathbf{u}_{mj}|\} \quad (29)$$

将每一层中的每一个子块的特征描述拼接起来得到最终的图像描述特征向量。

## 3 实验与分析

为验证本文提出的图像描述方法的有效性,选用了Caltech-101<sup>[32]</sup>和Caltech-256<sup>[33]</sup>图像数据库进行实验。首先将图像均转换为灰度图,然后通过DOF(Difference of Gaussian)方法检测关键特征点,关键点描述采用SIFT特征描述子,由局部 $16 \times 16$ 像素局部块概率密度采样得到的 $4 \times 4$ 共16个种子点来描述,每个关键点产生128个数据,移动步长为8像素。字典初始化由K-means方法获得,特征字典大小为 $128 \times 1024$ 。实验过程中分类器均采用LibSVM(Library for Support Vector Machine,可访问<http://www.csie.ntu.edu.tw/~cjlin/libsvm/>)支持向量机分类器。

Caltech-101图像库中包含了动物、花、车等101类图像,共9144幅图片;每一类图像中有31到800幅图片不等,绝大部分图像为 $300 \times 300$ 像素。分别在每类图像中选取部分图像为训练集,其他的为测试集进行实验。表1为本文方法LRC-DT与其他相关方法在Caltech-101上的实验结果对比。其中,LRC-DT的特征字典采用低秩约束优化得到。由实验结果可以看出,LRC-DT字典训练方法有效地提高了图像分类的准确率,随着每类中训练样本数目增多,加入低秩约束可以更好地调整特征字典的基向量,达到更好的分类效果,但是当每类训练样本太少时则提升较小或有较小的降幅。

为了评价采用低秩优化特征字典前后的实验效果,将Lazebnik等<sup>[11]</sup>提出的SPM方法和文献[21]中ScSPM编码方法中的特征字典替换为加入LRC低秩约束的特征字典,并进行分类效果比较。图1为实验结果,从图中可以看出,在其他步骤计算方法不变的情况下,通过将特征字典构造方法替换

为低秩约束方法,分类准确率均有不同程度的提升。

表1 Caltech-101 上图像分类准确率 %

每类参与训练图像个数	Zhang <sup>[34]</sup>	Lazebnik <sup>[11]</sup>	Yang <sup>[21]</sup>	Wang <sup>[22]</sup>	LRC-DT
5	46.6	43.20	56.54	51.15	52.60
10	55.8	51.96	59.45	59.77	61.60
15	59.1	56.40	67.00	65.43	68.73
20	62.0	58.75	68.70	67.74	70.05
25	—	62.10	70.15	70.16	72.34
30	66.2	64.60	73.20	73.44	76.10

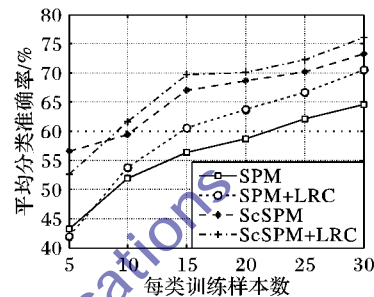


图1 Caltech-101 库上的结果比较

Caltech-256图像库有256类图像,共30607张图片,相比Caltech-101库有更多的类别和图像的变化,每类有至少80张图片。实验结果如表2所示,当训练样本个数为45和60,即每类中训练样本较多时,本文提出的方法分类准确率是最高的。

表2 Caltech-256 上图像分类准确率 %

每类参与训练图像个数	Griffin <sup>[35]</sup>	Yang <sup>[21]</sup>	Wang <sup>[22]</sup>	LRC-DT
15	28.3	27.73	34.36	33.15
30	34.1	34.02	41.19	39.76
45	—	37.46	45.31	46.17
60	—	40.14	47.68	49.78

图2所示为线性核函数下SPM方法与LLC编码方法中将特征字典替换为加入低秩优化约束特征字典的对比实验结果,可以看出加入LRC后准确率仍有部分提升。

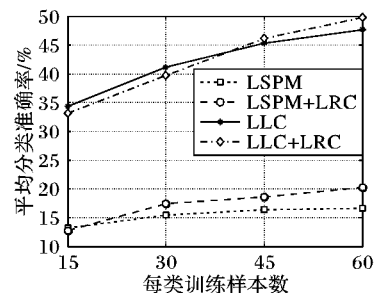


图2 Caltech-256 库上的结果比较

## 4 结语

构造表达能力丰富的多字典模型至关重要,而好的特征字典模型不仅可以表达丰富的图像特征,而且对不同类别图像的描述具有较强的可分辨性,同时对同类图像的描述具有相似性。本文提出的低秩约束同类图像系数矩阵的方法,将类别信息融入了特征字典的创建过程中,提高了特征字典的性能,一定程度上提高了基于字典模型的图像分类的准确率。

本文方法将同类图像特征描述相似的特性融入到了特征字典的构造中去,然而如何将异类图像间特征描述差异较大的特性加入进来,将在后续工作中进行研究;此外由于低秩优化计算比较耗时,低秩优化的高效计算也是后续工作的一个研究方向。

#### 参考文献:

- [1] JEGOU H, DOUZE M, SCHMID C. Packing bag-of-features [C]// Proceedings of the IEEE 12th International Conference on Computer Vision. Piscataway: IEEE, 2009: 2357–2364.
- [2] PEYRE G. A review of adaptive image representations [J]. IEEE Journal of Selected Topics in Signal Processing, 2011, 5(5): 896–911.
- [3] PHILBIN J, CHUM O, ISARD M, *et al.* Object retrieval with large vocabularies and fast spatial matching [C]// Proceedings of the 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2007: 1–8.
- [4] CHUM O, PHILBIN J, SIVIC J, *et al.* Total recall: automatic query expansion with a generative feature model for object retrieval [C]// Proceedings of the IEEE 11th International Conference on Computer Vision. Piscataway: IEEE, 2007: 1–8.
- [5] LI F F, PERONA P. A Bayesian hierarchical model for learning natural scene categories [C]// Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2005, 2: 524–531.
- [6] DENG C, CAO H. Construction of multiscale ridgelet dictionary and its application for image coding [J]. Journal of Image and Graphics, 2009, 14(7): 1273–1278. (邓承志, 曹汉强. 多尺度脊波字典的构造及其在图像编码中的应用[J]. 中国图象图形学报, 2009, 14(7): 1273–1278.)
- [7] SUN Y, WEI Z, XIAO L, *et al.* Multimorphology sparsity regularized image super-resolution [J]. Acta Electronica Sinica, 2010, 38(12): 2898–2903. (孙玉宝, 韦志辉, 肖亮, 等. 多形态稀疏性正则化的图像超分辨率算法[J]. 电子学报, 2010, 38(12): 2898–2903.)
- [8] AHARON M, ELAD M, BRUCKSTEIN A. The  $K$ -SVD: an algorithm for designing of overcomplete dictionaries for sparse representation [J]. IEEE Transactions on Signal Processing, 2006, 54(11): 4311–4322.
- [9] CSURKA G, DANCE C R, FAN L, *et al.* Visual categorization with bags of keypoints [C]// ECCV 2004: Proceedings of the 8th European Conference on Computer Vision, LNCS 3024. Berlin: Springer-Verlag, 2004: 1–22.
- [10] CSURKA G, DANCE C R, PERRONNIN F, *et al.* Generic visual categorization using weak geometry [C]// Toward Category-Level Object Recognition, LNCS 4170. Berlin: Springer-Verlag, 2006: 207–224.
- [11] LAZEBNIK S, SCHMID C, PONCE J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories [C]// Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2006, 2: 2169–2176.
- [12] ZHANG J, MARSZALEK M, LAZEBNIK S, *et al.* Local features and kernels for classification of texture and object categories: a comprehensive study [J]. International Journal of Computer Vision, 2007, 73(2): 213–238.
- [13] SIVIC J, ZISSERMAN A. Video Google: a text retrieval approach to object matching in videos [C]// Proceedings of the 9th IEEE International Conference on Computer Vision. Washington, DC: IEEE Computer Society, 2002: 1470–1477.
- [14] BOIMAN O, SHECHTMAN E, IRANI M. In defense of nearest-neighbor based image classification [C]// Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2008: 1–8.
- [15] LI F F, FERGUS R, PERONA P. Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories [J]. Computer Vision and Image Understanding, 2007, 106(1): 59–70.
- [16] BOSCH A, ZISSERMAN A, MUNOZ X. Scene classification using a hybrid generative/ discriminative approach [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(4): 712–727.
- [17] LIU Y, JIN R, SUKTHANKAR R, *et al.* Unifying discriminative visual codebook generation with classifier training for object category recognition [C]// Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2008: 1–8.
- [18] JURIE F, TRIGGS B. Creating efficient codebooks for visual recognition [C]// ICCV 2005: Proceedings of the 2005 IEEE International Conference on Computer Vision. Washington, DC: IEEE Computer Society, 2005, 1: 604–610.
- [19] BURGHOUTS G J, SCHUTTE K. Spatio-temporal layout of human actions for improved bag-of-words action detection [J]. Pattern Recognition Letters, 2013, 34(15): 1861–1869.
- [20] BANERJI S, SINHA A, LIU C. A new Bag of Words LBP (BoWL) descriptor for scene image classification [C]// CAIP 2013: Proceedings of the 15th International Conference on Computer Analysis of Images and Patterns, LNCS 8047. Berlin: Springer-Verlag, 2013: 490–497.
- [21] YANG J, YU K, GONG Y, *et al.* Linear spatial pyramid matching using sparse coding for image classification [C]// Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2009: 1794–1801.
- [22] WANG J, YANG J, YU K, *et al.* Locality-constrained linear coding for image classification [C]// Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2010: 3360–3367.
- [23] RUBINSTEIN, R, PELEG T, *et al.* Analysis  $K$ -SVD: a dictionary-learning algorithm for the analysis sparse model [J]. IEEE Transactions on Signal Processing, 2013, 61(3): 661–677.
- [24] JIANG Z, LIN Z, DAVIS L S. Label consistent  $K$ -SVD: learning a discriminative dictionary for recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(11): 2651–2664.
- [25] ZHANG Z, GANESH A, LIANG X, *et al.* TILT: transform invariant low-rank textures [J]. International Journal of Computer Vision, 2012, 99(1): 1–24.
- [26] LIU G, LIN Z, YAN C, *et al.* Robust recovery of subspace structures by low-rank representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 171–184.

表1 文献[16]与本文方法在拼接过程中内存消耗比较 MB

源图 像数	全局全景图 拼接方法	序列全景图 拼接方法	源图 像数	全局全景图 拼接方法	序列全景图 拼接方法
2	12.3	9.9	6	18.0	13.2
3	13.2	10.6	7	20.3	13.5
4	15.2	11.0	8	21.4	13.7
5	16.5	12.6	9	23.6	14.5

### 3 结语

本文提出了一种适用于移动设备的全景图拼接算法:一方面改进了传统图像拼接流程,即在图像拼接前先进行色彩校正,不仅改善了图像质量,而且减少了在消除接缝时泊松方程解算时间,同时采用梯度域移除“鬼影”的方法能较彻底地消除“鬼影”;另一方面,优化了图像拼接时内存分配机制。以往研究在图像拼接时,通常将所有源图像序列载入内存再进行拼接,而本文采用只需将当前复合图像和源图像载入内存,减少了内存消耗。

在不同光照环境下进行实验并与传统算法进行对比,从结果来看,本文算法能较好地解决全景图接缝及“鬼影”问题,拼接速度较快,同时占用较少的系统资源,图像质量比较令人满意,适合移动设备上高质量全景图的生成,具有一定的实用价值。

#### 参考文献:

- [1] PRABAKARAN G, BHAVANI R. A modified secure digital image steganography based on discrete wavelet transform [C]// ICCEET 2012: Proceedings of the 2012 International Conference on Computing, Electronics and Electrical Technologies. Piscataway: IEEE, 2012: 1096–1100.
- [2] WEN H, ZHOU J. An improved algorithm for image mosaic [C]// ISISE'08: Proceedings of the 2008 International Symposium on Information Science and Engineering. Piscataway: IEEE, 2008: 497–500.
- [3] JIAYA J, TANG C. Eliminating structure and intensity misalignment in image stitching [C]// ICCV 2005: Proceedings of the 2005 International Conference on Computer Vision. Piscataway: IEEE, 2005: 1651–1658.
- [4] WAN Y, MIAO Z. Automatic panorama image mosaic and ghost eliminating [C]// ICME 2008: Proceedings of the 2008 International Conference on Multimedia and Expo. Piscataway: IEEE, 2008: 945–948.
- [5] UYTENDAELE M, EDEN A, SZELISKI R. Eliminating ghosting and exposure artifacts in image mosaics [C]// CVPR 2001: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2001: 509–516.
- [6] DAVIS J. Mosaics of scenes with moving objects [C]// CVPR 1998: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 1998: 354–360.
- [7] YU T, HUI Y. Highly efficient image stitching based on energy map [C]// CISP 2009: Proceedings of the 2009 International Congress on Image and Signal Processing. Piscataway: IEEE, 2009: 1–5.
- [8] TAN K. Research on image mosaic technology [D]. Nanjing: Nanjing University of Science and Technology, 2006. (谭康. 图像拼接技术研究[ D]. 南京: 南京理工大学, 2006.)
- [9] FENG Y. Research on fast image registration and automatic mosaic [D]. Beijing: University of Chinese Academy of Sciences, 2010. (冯宇平. 图像快速配准与自动拼接技术研究[ D]. 北京: 中国科学院大学, 2010.)
- [10] RUAN P. Research on multi-image mosaic algorithm [D]. Changsha: Central South University, 2009. (阮鹏. 多图像拼接算法研究[ D]. 长沙: 中南大学, 2009.)
- [11] XIONG Y, PULLI K. Gradient domain image blending and implementation on mobile devices [C]// MobiCASE 2009: Proceedings of the 1st International ICST Conference on Mobile Computing, Applications and Services. Berlin: Springer-Verlag, 2010: 293–306.
- [12] HA S, KOO H, LEE S, *et al.* Panorama mosaic optimization for mobile camera systems [J]. IEEE Transactions on Consumer Electronics, 2007, 53(4): 1217–1225.
- [13] HA S, LEE S, CHO N, *et al.* Embedded panoramic mosaic system using auto-shot interface [J]. IEEE Transactions on Consumer Electronics, 2008, 54(1): 16–24.
- [14] XIONG Y. Eliminating ghosting artifacts for panoramic images [C]// ISM 2009: Proceedings of the 11st IEEE Multimedia Conference. Piscataway: IEEE, 2009: 432–437.
- [15] XIONG Y, PULLI K. Fast panorama stitching for high-quality panoramic images on mobile phones [J]. IEEE Transactions on Consumer Electronics, 2010, 56(2): 298–306.
- [16] CRIMINISI A, PEREZ P, TOYAMA K. Region filling and object removal by exemplar-based image inpainting [J]. IEEE Transactions on Image Processing, 2003, 13(9): 721–727.
- [17] (上接第 2672 页)
- [27] ZHANG N, YANG J. Low-rank representation based discriminative projection for robust feature extraction [J]. Neurocomputing, 2013, 111: 13–20.
- [28] SHALIT U, WEINSHALL D, CHECHIK G. Online learning in the embedded manifold of low-rank matrices [J]. Journal of Machine Learning Research, 2012, 13(1): 429–458.
- [29] LIU Y, JIAO L C, SHANG F, *et al.* An efficient matrix bi-factorization alternative optimization method for low-rank matrix recovery and completion [J]. Neural Networks, 2013, 48: 8–18.
- [30] ZHANG X, SUN F, LIU G, *et al.* Fast low-rank subspace segmentation [J]. IEEE Transactions on Knowledge and Data Engineering, 2013, 26(5): 1293–1297.
- [31] YANG J, YIN W T, ZHANG Y, *et al.* A fast algorithm for edge preserving variational multichannel image restoration [J]. SIAM Journal on Imaging Sciences, 2009, 2(2): 569–592.
- [32] BOIMAN O, SHECHTMAN E, IRANI M. In defense of nearest-neighbor based image classification [C]// CVPR 2008: Proceedings of the 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2008: 1–8.
- [33] SHOTTON J, WINN J, ROTHER C, *et al.* Textonboost for image understanding: multi-class object recognition and segmentation by jointly modeling appearance, shape and context [J]. International Journal of Computer Vision, 2009, 81(1): 2–23.
- [34] ZHANG H, BERG A C, MAIRE M, *et al.* SVM-KNN: discriminative nearest neighbor classification for visual category recognition [C]// Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2006, 2: 2126–2136.
- [35] GRIFFIN G, HOLUB A, PERONA P. Caltech-256 object category dataset, TR 7694 [R]. Pasadena: California Institute of Technology, 2007.