

基于非负弹性网稀疏编码算法的图像分类方法

张 勇^{1a,1b,2}, 张阳阳^{1a}, 程 洪^{1b}, 张艳霞^{1a}

(1. 电子科技大学 a. 数学科学学院; b. 自动化工程学院, 成都 611731; 2. 大数据研究中心, 成都 611731)

摘 要: 为提高图像分类的准确率, 提出一种非负弹性网稀疏编码算法。利用非负稀疏编码算法和弹性网模型, 在稀疏编码优化模型的目标函数中引入 l_2 范数正则项, 增加编码系数的非负约束, 并将该算法与空间金字塔模型相结合应用于图像分类。实验结果表明, 与传统的稀疏编码算法相比, 该算法不仅能提高编码的判别性与有效性, 而且可使相似的特征描述符编码后仍然相似, 增强编码的稳定性, 具有较高的分类准确度。

关键词: 图像分类; 稀疏编码; 空间金字塔匹配; 弹性网; 字典学习; 支持向量机

中文引用格式: 张 勇, 张阳阳, 程 洪, 等. 基于非负弹性网稀疏编码算法的图像分类方法[J]. 计算机工程, 2017, 43(7): 239-243, 249.

英文引用格式: Zhang Yong, Zhang Yangyang, Cheng Hong, et al. Image Classification Method Based on Non-negative Elastic Net Sparse Coding Algorithm[J]. Computer Engineering, 2017, 43(7): 239-243, 249.

Image Classification Method Based on Non-negative Elastic Net Sparse Coding Algorithm

ZHANG Yong^{1a,1b,2}, ZHANG Yangyang^{1a}, CHENG Hong^{1b}, ZHANG Yanxia^{1a}

(1a. School of Mathematical Science; 1b. School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China; 2. Big Data Research Center, Chengdu 611731, China)

[Abstract] In order to improve the image classification accuracy, this paper proposes a Non-negative Elastic Net Sparse Coding (NENSC) algorithm. This algorithm combines the advantages of non-negative sparse coding and elastic net algorithm. It introduces an l_2 norm regularization term to the objective function of Sparse Coding (SC) optimization model and non-negative constraints to coding coefficients are applied. The proposed algorithm combined with Spatial Pyramid Matching (SPM) model is applied to image classification. Experimental results show that, compared with the traditional sparse coding algorithm, the proposed algorithm not only increases the prediction capability and effectiveness of the coding, but also makes the similar feature descriptors similar after coding and improves the stability of the coding, it has higher classification accuracy.

[Key words] image classification; Sparse Coding (SC); Spatial Pyramid Matching (SPM); elastic net; dictionary learning; Support Vector Machine (SVM)

DOI: 10.3969/j.issn.1000-3428.2017.07.040

0 概述

随着智能设备和互联网技术的高速发展, 每天都有大量的图片通过各种渠道出现, 面对如此海量的图片数据, 要快速准确地找到自己感兴趣的图片变得越来越困难。因此, 图像分类技术引起了研究者的广泛关注, 成为模式识别和计算机视觉领域最重要的研究内容之一。

在众多的图像分类模型中, 以词袋 (Bag-of-

feature, BOF) 模型^[1]应用最为广泛, 它是从文本检索领域的 BOW 模型演变而来的, 即将图像表示为无序特征的集合。BOF 模型通常使用 K-means 算法将局部特征聚类, 形成视觉词典, 然后图像中的每个特征将被映射到视觉词典的某个词汇上, 每幅图像可以由视觉词汇的直方图表示。BOF 模型设计理念和计算简单易操作, 并且对图像的特征平移具有较强的鲁棒性, 但是却忽略了局部特征描述符之间的空间结构关系, 为此, 文献[2]提出了空间金字塔匹配

基金项目: 国家自然科学基金(11271001); 中央高校基本科研业务费专项资金(ZYGX2014Z012)。

作者简介: 张 勇(1974—), 男, 副教授、博士, 主研方向为图像处理、机器学习; 张阳阳, 硕士研究生; 程 洪, 教授; 张艳霞, 硕士。

收稿日期: 2016-06-28 **修回日期:** 2016-08-10 **E-mail:** mathzy@163.com

(Spatial Pyramid Matching, SPM)模型,将图像逐层分割成更加精细的空间子区域,然后分别计算其子区域中的局部特征直方图,在基准数据集 Caltech-101 和 Caltech-256 上取得了显著的分类效果。然而,SPM 模型在编码阶段通常利用向量量化(Vector Quantization, VQ)算法,每个特征描述子通过最近邻算法找到与其距离最近的视觉词汇,使得编码系数有且只有一个非零元素,并且为了得到较好的分类效果,通常需要利用非线性核函数进行分类,因此计算复杂度非常高。针对这些问题,文献[3]提出了稀疏编码和线性核相结合的 ScSPM 图像分类方法,用稀疏编码(Sparse Coding, SC)代替 VQ。ScSPM 减小了其重构误差,并且大大减少了计算复杂度,分类性能得到了进一步提升。

稀疏编码在图像分类中得到了广泛的应用^[4-7],传统的稀疏编码算法对特征描述符的变化非常敏感,原本相近的特征描述符在编码过程中可能由完全不同的视觉词汇表示。文献[5-7]对其进行改进,文献[5]提出了 LLC 编码算法,将局部信息加入原始的稀疏编码算法中,使得相近的特征描述符通过其局部近邻视觉词汇,得到相似的编码系数。文献[6]提出了快速近似局部线性编码方法,并通过增量训练构建 LLC 码本,然后利用其将每个描述符映射到本地坐标中,快速准确地表示了特征描述符之间的相似性。文献[7]提出了基于弹性网规则的字典学习方法,不仅可以产生与稀疏编码算法相似的稀疏系数,还可以有效地控制特征描述符的敏感性。但这些方法应用框架较为固定,较难进行推广。

ScSPM 在编码过程中,为了最小化特征的重构误差,得到的编码系数可能会有负数元素出现,加上后面的最大池化过程,负数元素最终被零取代,极易造成图像信息的丢失,最终影响图像的分类准确率。如何得到非负编码系数,成为近年来一个重要的研究热点。文献[8]通过在 LLC 优化模型的目标函数中加入非负约束,提出了非负局部约束线性编码,有效地提高了模型的稳定性。文献[9]利用非负稀疏编码、低秩与稀疏矩阵分解方法构建的图像分类模型,在一些基准数据集上取得了较好的效果。

为有效克服传统 l_1 范数约束下稀疏编码存在的敏感性和信息丢失问题,提高编码的判别性和有效性,本文提出一种非负弹性网稀疏编码(Non-Negative Elastic Net Sparse Coding, NENSC)算法。

1 基于 SPM 模型的图像分类

传统的基于 SPM 模型的图像分类主要分为4个

步骤,即图像特征提取、特征编码、SPM 分块及 Pooling、SVM 分类。首先,从输入图像中提取特征,常用方法有 SIFT, LBP, HOG 等;其次,对提取的每个特征描述子进行编码形成一个 D 维的编码向量;然后,用 SPM 方法对图像在 2 个坐标方向上进行 $2^L \times 2^L$ 次划分, L 取 0, 1, 2, 将每层的各个子区域池化,形成 M 维编码向量,则第 L 层会得到 $2^L \times 2^L$ 个 M 维的编码向量,将该层的所有编码向量拼接,得到该层的编码向量,再将三层的编码向量拼接,得到一个 $21 \times M$ 维的编码向量,即是图像的最终向量表示;最后,将该向量输入到分类器中进行分类。当 $L=0$ 时即为 BOF 模型。

编码阶段隐含了视觉词典的构建过程,通常是采用 K-means 算法对全体特征描述符聚类得到视觉词典。假设 $X \in \mathbb{R}^{D \times N}$ 是从输入图像中由 SIFT 提取出来的局部特征描述子, $X = [x_1, x_2, \dots, x_N]$,

$B \in \mathbb{R}^{D \times M}$ 是视觉词典,不同的编码方式将每一个描述符映射成不同的 M 维向量,最终形成不同的图像表示向量。图像分类的整体框架如图 1 所示。

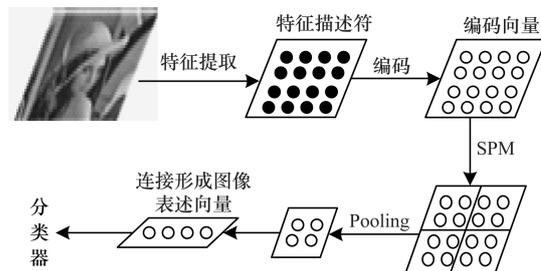


图 1 基于 SPM 模型的图像分类整体框架

1.1 向量量化

向量量化(VQ)方法是最原始的编码方法,研究者通常利用 VQ 来解决下列优化问题:

$$\operatorname{argmin}_B \sum_{i=1}^N \min_{j=1,2,\dots,M} \|x_i - b_j\|_2^2 \quad (1)$$

其中, $B = [b_1, b_2, \dots, b_M]$ 是视觉词典,又称为码本, b_i 是视觉单词(Visual Word); $\|\cdot\|_2$ 是向量的 l_2 范数。VQ 的基本思想在于在视觉词典 B 中寻找一个距离原始特征 x_i 最近的视觉单词来重新表示该原始特征。式(1)中的优化问题可以转化为下面的最小二乘问题:

$$\operatorname{argmin}_Y \sum_{i=1}^N \|x_i - By_i\|_2^2 \quad (2)$$

s. t. $\|y_i\|_0 = 1, \|y_i\|_1 = 1, y_i \geq 0, \forall i$

其中, $Y = [y_1, y_2, \dots, y_N]$ 是特征 X 对应的编码集。 $\|y_i\|_0 = 1, \|y_i\|_1 = 1, y_i \geq 0$ 分别代表每一个 y_i 只有一个非零元素, y_i 中的元素都大于 0, 且所有元素

之和等于 1。可以看出,在 VQ 方法中每一个特征只能分配给距离它最近的一个视觉词汇上,并且与之相关的权重固定为 1。

1.2 基于 l_1 范数的稀疏编码

为了放松向量量化中对视觉词汇的约束,文献[3]提出了 ScSPM 方法,即将编码系数 y_i 的 l_1 范数正则项加入到式(2)中,使得特征 x_i 可以由多个视觉词汇表示。相比于向量量化,ScSPM 在几种常用的基准数据集上取得了非常好的实验效果。其优化目标函数为:

$$\begin{aligned} \operatorname{argmin}_{\mathbf{B}, \mathbf{Y}} \sum_{i=1}^N \| \mathbf{x}_i - \mathbf{B} \mathbf{y}_i \|_2^2 + \lambda \| \mathbf{y}_i \|_1 \\ \text{s. t. } \| \mathbf{b}_j \|_2 \leq 1, \forall j = 1, 2, \dots, M \end{aligned} \quad (3)$$

与 VQ 的优化函数式(2)相比,式(3)是对视觉词典和编码系数的双优化问题。稀疏编码通常分为训练和测试 2 个阶段,在训练阶段求解这个优化问题,通常采用交替优化的方法,先固定其中一个,然后优化另一个,最终得到最优的字典 \mathbf{B} 和编码。训练阶段得到最优字典 \mathbf{B} 后,在测试阶段只需将待测试的特征进行编码即可,即在此阶段仅仅只需要优化 \mathbf{Y} 。

2 非负弹性网稀疏编码

基于 l_1 范数的稀疏编码在图像分类中取得了非常好的效果,但是研究者发现,该模型具有不稳定性,在编码过程中原本相近的特征描述符得到的编码向量可能差别很大。并且为了最小化模型的重构误差,编码向量中可能会有负数出现,加上后面的最大池化,负数元素最终被零取代,造成图像信息丢失,最终影响分类效果。文献[10]提出了非负稀疏编码模型,如式(4)所示。

$$\begin{aligned} \operatorname{argmin}_{\mathbf{B}, \mathbf{Y}} \sum_{i=1}^N \| \mathbf{x}_i - \mathbf{B} \mathbf{y}_i \|_2^2 + \lambda \| \mathbf{y}_i \|_1 \\ \text{s. t. } \| \mathbf{b}_j \|_2 \leq 1, |y_i| = 1, \mathbf{y}_i \geq 0, \forall i, j \end{aligned} \quad (4)$$

文献[11]提出了弹性网模型,具有很好的特征选择能力,不仅可以产生与稀疏编码模型相似的稀疏系数,还可以将高度相关的特征一起选出来,有效地控制了特征描述符的敏感性。弹性网模型如式(5)所示。

$$\begin{aligned} \operatorname{argmin}_{\mathbf{B}, \mathbf{Y}} \sum_{i=1}^N \| \mathbf{x}_i - \mathbf{B} \mathbf{y}_i \|_2^2 + \lambda \| \mathbf{y}_i \|_1 + \gamma \| \mathbf{y}_i \|_2^2 \\ \text{s. t. } \| \mathbf{b}_j \|_2 = 1, \mathbf{b}_j \geq 0, \forall i, j \end{aligned} \quad (5)$$

本文将非负稀疏编码和弹性网模型相结合,提出一种非负弹性网稀疏编码模型(NENSC)。

2.1 模型介绍

本文基于式(4)、式(5)的非负弹性网稀疏编码

模型描述如下:

$$\begin{aligned} \operatorname{argmin}_{\mathbf{B}, \mathbf{Y}} \sum_{i=1}^N \| \mathbf{x}_i - \mathbf{B} \mathbf{y}_i \|_2^2 + \lambda \| \mathbf{y}_i \|_1 + \gamma \| \mathbf{y}_i \|_2^2 \\ \text{s. t. } \mathbf{y}_i \geq 0, \| \mathbf{b}_j \|_2 = 1, \mathbf{b}_j \geq 0, \forall i, j \end{aligned} \quad (6)$$

其中,式(6)中的第 1 项称为重构误差项;第 2 项用于控制编码系数 y_i 的稀疏性, l_2 范数正则项用来控制模型的敏感性。 λ 和 γ 都是非负正则参数,当 γ 为 0 时,即为 ScSPM 模型;当 λ 为 0 时,式(6)即是著名的岭回归(Ridge Regression)。

2.2 模型的优化实现

$$\text{定义 } \bar{\mathbf{B}}_{(D+M) \times M} = (1 + \gamma)^{-\frac{1}{2}} \begin{pmatrix} \mathbf{B} \\ \sqrt{\gamma} \mathbf{I} \end{pmatrix}, \bar{\mathbf{x}}_{D+M} = \begin{pmatrix} \mathbf{x} \\ 0 \end{pmatrix},$$

并且令^[11]:

$$\beta = \frac{\lambda}{\sqrt{1 + \gamma}}, \bar{\mathbf{y}} = \sqrt{1 + \gamma} \mathbf{y} \quad (7)$$

可以转化为:

$$\operatorname{argmin}_{\bar{\mathbf{B}}, \bar{\mathbf{Y}}} \sum_{i=1}^N \| \bar{\mathbf{x}}_i - \bar{\mathbf{B}} \bar{\mathbf{y}}_i \|_2^2 + \beta \| \bar{\mathbf{y}}_i \| \quad (8)$$

2.2.1 字典学习

式(8)所述的优化问题,对于 $\bar{\mathbf{B}}$ 和 $\bar{\mathbf{Y}}$ 是非凸的,但是当固定其中一个求解另一个时是凸的。固定 $\bar{\mathbf{Y}}$,则可以转化为求解下列问题:

$$\operatorname{argmin}_{\bar{\mathbf{B}}} \sum_{i=1}^N \| \bar{\mathbf{x}}_i - \bar{\mathbf{B}} \bar{\mathbf{y}}_i \|_2^2 \quad (9)$$

此优化问题通常可以利用梯度下降法求解,这里利用拉格朗日对偶算法求解更加方便。假设 λ 是对偶变量,则拉格朗日函数可以表述为:

$$\begin{aligned} L(\bar{\mathbf{B}}, \lambda) = \operatorname{tr}((\bar{\mathbf{X}} - \bar{\mathbf{B}} \bar{\mathbf{Y}})^T (\bar{\mathbf{X}} - \bar{\mathbf{B}} \bar{\mathbf{Y}})) \\ + \sum_{j=1}^D \lambda_j \left(\sum_{i=1}^M \bar{\mathbf{B}}_{i,j}^2 - 1 \right) \end{aligned}$$

其中, $\operatorname{tr}(A) = \operatorname{trace}(A)$, 则可得到拉格朗日对偶函数如下:

$$\begin{aligned} D(\lambda) = \operatorname{argmin}_{\bar{\mathbf{B}}} L(\bar{\mathbf{B}}, \lambda) \\ = \operatorname{tr}(\bar{\mathbf{X}} \bar{\mathbf{X}}^T - \bar{\mathbf{X}} \bar{\mathbf{Y}}^T (\bar{\mathbf{Y}} \bar{\mathbf{Y}}^T + \Lambda) - 1 (\bar{\mathbf{X}} \bar{\mathbf{Y}}^{-1}) - \Lambda) \end{aligned}$$

其中, $\Lambda = \operatorname{diag}(\lambda)$, $D(\lambda)$ 的梯度和 Hessian 阵计算如下:

$$\frac{\partial D(\lambda)}{\partial \lambda_i} = \| \bar{\mathbf{X}} \bar{\mathbf{Y}}^T (\bar{\mathbf{Y}} \bar{\mathbf{Y}}^T + \Lambda) \mathbf{e}_i \|^2 - 1$$

$$\frac{\partial D(\lambda)}{\partial \lambda_i \partial \lambda_j} = -2((\bar{\mathbf{Y}} \bar{\mathbf{Y}}^T + \Lambda)^{-1})_{i,j}$$

$$\times (\bar{\mathbf{X}} \bar{\mathbf{Y}})^T \bar{\mathbf{X}} \bar{\mathbf{Y}}^T (\bar{\mathbf{Y}} \bar{\mathbf{Y}}^T + \Lambda)^{-1}_{i,j} (\bar{\mathbf{Y}} \bar{\mathbf{Y}}^T + \Lambda)^{-1}_{i,j}$$

其中, $\mathbf{e}_i \in \mathbb{R}^n$ 是单位向量。此时再利用牛顿法得到上述朗格朗日对偶函数的解为:

$$\bar{\mathbf{B}}^T = (\bar{\mathbf{Y}} \bar{\mathbf{Y}}^T + \Lambda)^{-1} (\bar{\mathbf{X}} \bar{\mathbf{Y}}^T)^T \quad (10)$$

2.2.2 非负稀疏编码求解

对于式(8),若固定 $\bar{\mathbf{B}}$,则是一个系数带有非负约束的线性回归问题:

$$F(\bar{\mathbf{y}}) = \|\bar{\mathbf{x}} - \bar{\mathbf{B}}\bar{\mathbf{y}}\|_2^2 + \beta \|\bar{\mathbf{y}}\|_1 \quad (11)$$

依次计算 \mathbf{Y} 的每一列,优化方法如下:

$$\bar{\mathbf{y}} = \bar{\mathbf{y}}' * (\bar{\mathbf{B}}^T \bar{\mathbf{x}}) ./ (\bar{\mathbf{B}}^T \bar{\mathbf{B}} \bar{\mathbf{y}}' + 2\beta c)$$

其计算方法和证明过程与文献[10]类似。

非负弹性网稀疏编码算法的总体步骤下:

算法1 非负弹性网稀疏编码算法

输入 $\mathbf{X}, \lambda, \gamma$

输出 \mathbf{B}, \mathbf{Y}

步骤1 随机初始化 $\mathbf{B}^0, \mathbf{Y}^0$,其中 $B_{i,j} > 0, Y_{i,j} > 0$,同时使得 \mathbf{B}^0 的每一列 $\|\mathbf{b}_i\| = 1$ 。根据输入参数和上述定义求得 $\bar{\mathbf{X}}^0, \bar{\mathbf{B}}^0, \bar{\mathbf{Y}}^0$ 。

步骤2 重复以下步骤,直到满足收敛条件。

更新字典 \mathbf{B} :

步骤2.1

$$\bar{\mathbf{B}}^{(k)} \leftarrow ((\bar{\mathbf{Y}}(k) \bar{\mathbf{Y}}^T(k) + \Lambda)^{-1} (\bar{\mathbf{X}}(k) \bar{\mathbf{Y}}^T(k))^T)^T$$

其中, $\Lambda = \text{diag}(\lambda)$ 。

步骤2.2 将得到的字典 $\bar{\mathbf{B}}^{(k)}$ 的每个负数元素设为0,并使得其每一列 $\|\bar{\mathbf{b}}_i\| = 1, \bar{\mathbf{B}}^k = \bar{\mathbf{B}}^{(k)}$ 。

更新编码系数 \mathbf{Y} 的每一列:

$$\text{步骤2.3 } \bar{\mathbf{y}}^{k+1} = \bar{\mathbf{y}}^k \times ((\bar{\mathbf{B}}^k)^T \bar{\mathbf{x}}) ./ ((\bar{\mathbf{B}}^k)^T \times \bar{\mathbf{B}}^k \bar{\mathbf{y}}^{k+1} + 2\beta c)$$

步骤2.4 $k+1 \leftarrow k$

步骤3 $\mathbf{Y} = 1/\sqrt{1+\gamma} \bar{\mathbf{Y}}$

2.3 基于 NENSC 的图像分类模型

选取恰当的特征在图像分类中起着至关重要的作用。SIFT 是性能较好的局部特征描述子之一,其最大的优点是对旋转、尺度缩放、亮度变化等具有较强的稳定性,并且在有视觉变化和噪声影响的环境下对特征的提取也具有较好的鲁棒性。基于此,本文选择使用 SIFT 进行图像的特征提取。

本文将 NENSC 和 SPM 相结合运用到图像分类中,在对图像编码之后,将图像进行 $1 \times 1, 2 \times 2, 4 \times 4$ 的划分,构造3层空间金字塔,共有 $L=21$ 块子区域。对于每个子区域的编码向量合并形成一个向量,所有向量链接在一起形成最终的图像表示向量。因此,一幅图像的代表向量可以通过下面的函数求得:

$$\mathbf{I} = \{\mathbf{U}_1^T, \mathbf{U}_2^T, \dots, \mathbf{U}_L^T\}$$

其中,函数 $U_i = \max(Y_i^l, i \in M^l, l=1, 2, \dots, L), M^l$ 表示第 i 块子区域的特征。

基于 SC 和 NENSC 的图像分类框架如图2所示。

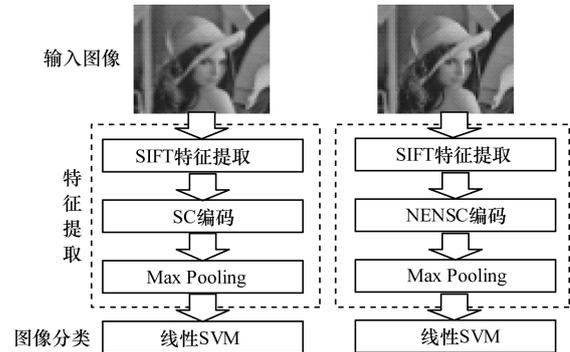


图2 基于 SC 和 NENSC 编码的图像分类框架

3 实验与结果分析

为了验证本文提出的非负弹性网稀疏编码算法的有效性,在4个常用的基准数据集 Caltech-101^[12], Scene-15^[13], UIUC-Sport^[14], Caltech-256^[15]上进行测试,并将各自的分类准确率与 KSPM, KC, ScSPM, 本文方法等进行比较分析^[16]。

3.1 算法复杂度分析

为了获得较高的分类准确率,传统的 BOW 和 SPM 模型必须使用非线性核函数,如 Mercer kernels 等。而非线性 SVM 在训练阶段所需的时间复杂度为 $O(n^3)$,这里 n 代表训练样本的个数;因为支持向量的个数是 n 的线性函数,所以在测试阶段的时间复杂度为 $O(n)$ 。本文方法与文献[3]提出的 ScSPM 一样使用线性 SVM,该方法在训练阶段可以将时间复杂度降低为 $O(n)$,在测试阶段的时间复杂度与 BOW 和 SPM 模型相同,并取得了比其更好的分类准确率。

3.2 数据集及实验设置

在 SIFT 提取特征时,按照步长为8像素密集采样。视觉词典的维数设置为1024维。对于 SVM,本文使用 LibSVM Package^[17]开发包。本文同文献[3]所用的 λ 保持一致,设置为0.3, γ 设置为0.06。每次实验重复进行10次,最后用平均值和方差作为最终的识别率。

1) Caltech-101 数据集:该数据集共有102类(包括动物、花草、交通工具等),共含有9144幅图像。每一类的图像数目为31~800不等。为了与其他模型对比,公平起见,根据之前研究者的实验设置,每类分别随机选取15幅和30幅图像进行训练,剩下的进行测试。

2) Caltech-256 数据集:该数据集共有 257 类,共含有 30 607 幅图像。每一类的图像数目不少于 80 幅。在实验中每类随机选取 60 幅图像进行训练,剩下的进行测试。

3) 15-Scenes 数据集:该数据集共有 15 类,包含 4 485 幅图像。每一类的图像数目为 200 ~ 400 不等。在实验中每类随机选取 100 幅图像进行训练,剩下的进行测试。

4) UIUC-Sport 数据集:该数据集共 8 类,1 792 幅图像,每一类的图像数目为 137 ~ 250 不等。在实验中每类随机选取 70 幅图像进行训练,剩下的进行测试。

3.3 实验结果

在表 1 中列举了本文所提模型在 Caltech-101 上的实验结果,并同时列举了当前比较流行的几种模型在相同实验条件的实验结果进行比较,尤其需要注意的是,对于 LLC^[5] 模型,使用文献[3]提供的源代码,并将字典大小设置为 1 024 维,在基准数据集上进行实验。当训练集为 30 时,共有 12 类取得了 100 % 的分类准确率。

表 1 Caltech-101 数据集上分类准确率对比 %

方法	15 training	30 training
KSPM 方法 ^[2]	56.44 ± 0.78	63.99 ± 0.88
KC 方法 ^[16]	-	64.14 ± 1.88
ScSPM ^[3]	67.0 ± 0.45	73.21 ± 0.54
LLC 方法 ^[5]	63.35 ± 0.96	70.20 ± 0.74
本文方法	67.31 ± 0.87	73.83 ± 0.81

在表 2 中列举了本文所提模型在 Caltech-256, 15-Scence 和 UIUC-Sport 上的实验结果。

表 2 3 个基准数据集上分类准确率对比 %

方法	Caltech-256	15-Scence	UIUC-Sport
KSPM 方法 ^[2]	-	81.40 ± 0.50	-
KC 方法 ^[16]	-	76.67 ± 0.39	-
ScSPM 方法 ^[3]	40.14 ± 0.91	80.28 ± 0.93	82.74 ± 1.46
本文方法	41.90 ± 0.52	82.51 ± 0.72	84.38 ± 0.98

从表 1、表 2 中可以看到,对于 15-Scene 数据集,在训练数据集和字典维数相同的情况下,本文提出的算法比传统的基于 l_1 范数的稀疏编码(ScSPM)方法分类准确率要高出约 2 %。在 Caltech-256 和 UIUC-Sport 数据集上,本文提出的算法高出 1 % ~ 2 %。

4 结束语

稀疏编码在图像分类中的应用受到越来越多的关注,本文对传统的基于 l_1 范数的稀疏编码算法存在的不足进行改进,提出一种非负弹性网稀疏编码算法,并将其应用于图像分类,有效地抑制了原模型

对图像信息丢失和对特征描述符敏感的缺点,提高了模型的稳定性,使得模型的泛化能力得到了增强。本文利用 4 个基准数据集来验证所提算法的有效性,实验结果表明,该算法在相同条件下图像的分类准确率优于传统的稀疏编码算法。

参考文献

- [1] Csurka G, Dance C R, Fan Lixin, et al. Visual Categorization with Bags of Keypoints [C]//Proceedings of European Conference Computer Vision. Prague, Czech Republic: IEEE Press, 2004: 59-74.
- [2] Lazebnik S, Schmid C, Ponce J. Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2006: 2169-2178.
- [3] Yang J, Yu K, Gong Y, et al. Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE Press, 2009: 1794-1801.
- [4] 李钱钱,曹 国. 基于拉普拉斯非负稀疏编码的图像分类[J]. 计算机工程, 2013, 39(11): 240-244.
- [5] Wang J, Yang J, Yu K, et al. Locality-constrained Linear Coding for Image Classification [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE Press, 2010: 3360-3367.
- [6] 王元强,郝尚富. 一种基于局部约束线性编码的图像分类方法[J]. 计算机工程, 2015, 41(7): 256-260.
- [7] Shen Bin, Liu Baodi, Wang Qifan. Elastic Net Regularized Dictionary Learning for Image Classification [J]. Multimedia Tools & Applications, 2016, 75(15): 1-14.
- [8] 刘培娜,刘国军,郭茂祖,等. 非负局部约束线性编码图像分类算法[J]. 自动化学报, 2015, 41(7): 1235-1243.
- [9] Zhang C J, Liu J, Tian Q, et al. Image Classification by Non-negative Sparse Coding, Low-rank and Sparse Decomposition [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Colorado, USA: IEEE. Press, 2011: 1673-1680.
- [10] Hoyer P O. Non-negative Sparse Coding [C]//Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing. Washington D. C., USA: IEEE Press, 2002: 557-565.
- [11] Zou H, Hastie T. Regularization and Variable Selection via the Elastic Net [J]. Journal of the Royal Statistical Society, 2005, 67(2): 301-320.
- [12] Li F F, Fergus R, Perona P. Learning Generative Visual Models from Few Training Examples: An Incremental Bayesian Approach Tested on 101 Object Categories [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE Press, 2004: 59-74.

2.3 深度图

获得视差图后,可以通过立体投影原理求得 Teddy 的深度图。本文实验得到的深度图效果如图 11 所示。



图 11 Teddy 深度图

3 结束语

本文提出一种改进的全局立体匹配算法。采用全局差错能量函数作为立体匹配的匹配代价,进行跳跃式区域生长,隔点求取差错能量函数从而获取视差图。通过对比文献[16]算法的视差图,并且经过实验验证,表明本文算法在保证准确性的基础上具有很好的实时性,能够节省 50% 以上的时间。下一步工作将研究如何获取更高精度的视差图以及对目标物体形态信息进行三维重建。

参考文献

- [1] 韩九强. 机器视觉技术及应用[M]. 北京:高等教育出版社,2009.
- [2] 周龙,徐贵力,李开宇,等. 无人机导航场景中立体匹配算法研究[J]. 航空兵器,2012(1):48-52.
- [3] Stentoumis C, Grammatikopoulos L, Kalisperakis I, et al. Alocal Adaptive Approach for Dense Stereo Matching in Architectural Scene Reconstruction [C]// Proceedings of the 5th International Workshop on 3D Virtual Reconstruction and Visualization of Complex Architectures. Trento, Italy: [s. n.], 2013.

- [4] 赵敏,刘秉琦,武东升. 增强现实装配和维修系统的技术研究[J]. 光学仪器,2012,34(2):16-20.
- [5] Yoon K J, Kweon I S. Locally Adaptive Support-weight Approach for Visual Correspondence Search [C]// Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington D. C., USA:IEEE Press,2005:924-931.
- [6] 卢迪,林雪. 多种相似性测度结合的局部立体匹配算法[J]. 机器人,2016,38(1):1-7.
- [7] 徐青,王敬东,李鹏,等. 基于图像分割的快速立体匹配算法[J]. 计算机工程,2006,32(22):209-211.
- [8] Khan W, Suaste V, Caudillo D. Belief Propagation Stereo Matching Compared to iSGM on Binocular or Trinocular Video Data [C]// Proceedings of IEEE Intelligent Vehicles Symposium. Washington D. C., USA: IEEE Press,2013:791-796.
- [9] Besse F, Rother C, Fitzgibbon A, et al. Patch Match Belief Propagation for Correspondence Field Estimation [J]. International Journal of Computer Vision, 2014, 110(1): 2-13.
- [10] 余科,谢红. 基于动态规划与图像分割的立体匹配算法[J]. 应用科技,2011,38(5):47-51.
- [11] Veksler O. Efficient Graph-based Energy Minimization Methods in Computer Vision [D]. New York, USA: Cornell University, 1999.
- [12] Banno A, Ikeuchi K. Disparity Map Refinement and 3D Surface Smoothing via Directed Anisotropic Diffusion [C]// Proceedings of the 12th IEEE International Conference on Computer Vision Workshops. Washington D. C., USA: IEEE Press, 2009:1870-1877.
- [13] 郁理,郭立,袁红星. 基于分级置信度传播的立体匹配新方法[J]. 中国图象图形学报,2011,16(1):103-109.
- [14] 于平,王士同. 半监督空间化竞争聚集算法及其在图像分割中的应用[J]. 计算机工程,2015,41(2):234-241.
- [15] 李娇. 基于立体视觉的双目匹配[D]. 南京:南京理工大学,2015.
- [16] Alagoz B B. Obtaining Depth Maps from Color Images by Region Based Stereo Matching Algorithms [J]. OncuBilim Algorithm and Systems Labs, 2008, 8(4).

编辑 金胡考

(上接第 243 页)

- [13] Li F F, Perona P. A Bayesian Hierarchical Model for Learning Natural Scene Categories [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Diego, USA: IEEE Press, 2005: 524-531.
- [14] Li J L, Li F F. What, where and who? Classifying Events by Scene and Object Recognition [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Riode, Brazil: IEEE Press, 2007:1-8.
- [15] Krizhevsky A, Sutskever I, Hinton G E. Imagenet Classification

with Deep Convolutional Neural Networks [C]// Proceeding of NIPS' 12. Washington D. C., USA: IEEE Press, 2012: 1097-1105.

- [16] Jan C, van G, Geusebroek J M, et al. Kernel Codebooks for Scene [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Marseille, France: IEEE Press, 2008:696-709.
- [17] Chang C C, Lin C J. LIBSVM: A Library for Support Vector Machines [J]. ACM Transactions on Intelligent Systems and Technology, 2011, 2(3):1-25.

编辑 索书志