doi:10.11887/j.cn.201402022 http://journal.nudt.edu.cn

基于梯度方向二进制模式的空间金字塔模型方法*

郭 军[1],周 晖[2],朱长仁[1],肖顺平[1]
(1.国防科技大学 电子科学与工程学院,湖南 长沙 410073;2.北京跟踪与通信技术研究所,北京 100094)

摘 要: 空间金字塔模型由于其优势在当前图像分类中得到了广泛应用。然而,其码本生成和特征量化这两个环节具有较高的计算复杂度。为了解决这个问题,提出了一种新的局部特征表述——梯度方向二进制模式,首先对图像稠密采样得到多个子图像块,再将每个子图像块均匀划分为2×2个网格,计算每个网格的梯度直方图,然后对所有网格的梯度主方向进行二进制编码并连接为二进制串值,该二进制串值转换的十进制数即为子图像块的特征表述,最后将该特征表述嵌入到SPM模型中。在标准分类数据库上的实验结果证明了本方法在算法耗时和分类精度上均优于基于SIFT的SPM方法。

关键词: 空间金字塔模型;梯度方向二进制模式;局部特征描述;图像分类
中图分类号: TP391 **文献标志码:** A **文章编号:** 1001-2486(2014)02-0129-05

A spatial pyramid model based on binary pattern of oriented gradients

GUO Jun[1], *ZHOU Hui*[2], *ZHU Changren*[1], *XIAO Shunping*[1]
(1. College of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China;
2. Beijing Institute of Tracking and Telecommunications Technology, Beijing 100094, China)

Abstract: Recently spatial pyramid matching (SPM) with scale invariant feature transform (SIFT) descriptor has been successfully used in image classification. Unfortunately, the codebook generation and feature quantization procedures using SIFT feature have the high complexity both in time and space. To address this problem, a feature descriptor called Binary Pattern of Oriented Gradients is presented. Firstly, the input image was densely sampled and divided into small uniform image patches. Secondly, each patch was divided into 2 * 2 grids uniformly. For all grids the histograms of oriented gradient were computed and all dominant directions of the histograms were coded by binary coding. Then the descriptor was generated by converting the binary number to decimal number. Finally, this descriptor was combined in the spatial pyramid domain. Experiments on popular benchmark dataset demonstrate that the proposed method always significantly outperforms the popular SPM based SIFT descriptor method both in time and classification accuracy.

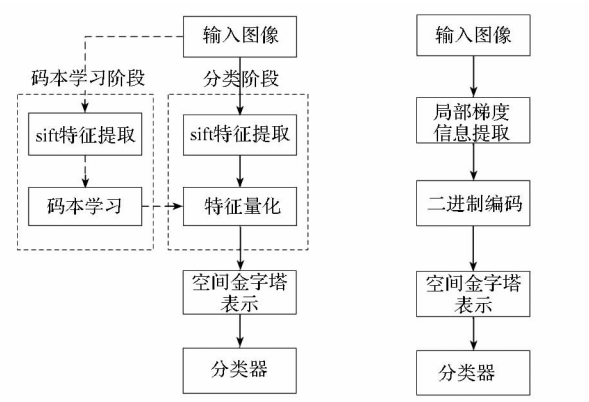
Key words: spatial pyramid model; binary pattern of oriented gradients; local feature descriptor; image classification

图像目标的分类是计算机视觉和模式识别领域的重要问题,图像特征的描述又是图像目标分类研究的一个重要方面。近年来,BOW(bag-of-words)方法[1-2]在图像特征表述中得到了非常广泛的应用。这一类方法首先对训练图像中的局部图像块提取出无序的表观特征描述,然后对特征描述的集合进行聚类操作,量化得到一些离散的代表性特征,称为视觉词汇(visual words)。将视觉词汇作为描述图像的基本元素,对每一幅图像统计出一个紧凑的直方图表述作为图像特征,再通过各种训练方法来学习图像的分类模型。BOW模型在图像分类、标注、检索和视频事件检测中显示出良好的性能。

BOW方法模型简单且效率高,但是却忽略了特征之间的空间关系,因此这种描述能力是受限制的。为了利用局部特征在图像空间的位置关系,Lazebnik等[3]提出BOW模型的一种扩展方法——空间金字塔匹配(Spatial Pyramid Matching,SPM)模型,该模型首先对图像划分稠密的图像块,对图像块提取SIFT特征描述,再对特征进行量化得到词汇表示后,在二维图像空间建立多级金字塔,然后计算加权的子图像区域局部特征直方图交叉。该方法在多个图像分类任务中体现出优异的性能。

* **收稿日期:** 2013-10-23
基金项目: 国家部委资助项目
作者简介: 郭军(1982—),男,新疆伊犁人,博士研究生,E-mail:guojun@nudt.edu.cn;
朱长仁(通信作者),男,副教授,博士,硕士生导师,E-mail:changrenzhu@nudt.edu.cn

SIFT 算子与 SPM 模型结合融入了对空间信息的刻画,表现出了很好的图像分类性能,但其继承了传统 BOW 模型中码本学习这个最为关键的环节,而这个环节本身存在着不足。在码本学习环节,码本的码长度是自定义的,根据类别的多少、分类的对象和训练样本的差别,码本长度影响着码本的描述能力,因此需要确定一个合适的长度。而在特征量化时,通常使用了基于最近邻匹配的向量量化,对图像中每一个待量化特征选取离其最近邻的一个视觉词汇作为其特征表示,由于用于码本学习的样本不可避免地包含着噪声,这也将对码本的描述能力造成影响。同时,码本学习和通过此码本映射图像的特征描述都具有一定的时间与空间复杂度。



(a)SPM 流程框架 (b)本文方法流程

图 1 本文方法与 SPM 模型流程图对比

Fig. 1 Schematic comparison of the original SPM with our proposed method

针对特征量化过程造成的误差问题,近年来许多学者在 SPM 模型的基础上提出了一些扩展模型[4-11]。Gemert 等[4]提出了一种模糊量化的方法,利用码本中的 K - 近邻词汇进行自适应线性加权来表示待量化的特征,但是对不同分类集选择合适的参数将对结果造成影响。Yang 等[5]使用稀疏编码模型来量化图像特征,通过解凸优化问题来求取待量化特征关于基向量的稀疏表示,使用多个向量的线性组合将图像特征表示为稀疏向量,减少了量化造成的误差,取得了较好的分类效果,虽然稀疏编码和 SPM 的结合使性能有所提升,但依然具有较高的计算复杂度。

此外,学者们在金字塔特征的汇总方面和金字塔网格划分方面做了深入研究。文献[6 - 7]将 SPM 方法[3]使用的特征平均汇总(average pooling)替换为特征各维最大汇总(max pooling),对不同尺度上的特征分别汇总后,串接为表述图像的全局向量。文献[10 - 11]则突破了 SPM 方法[3]分层规则划分金字塔网格的方式,将图像以不同尺寸的矩形网格过完备地划分为大量的子区域,通过对这些子区域学习训练生成码本词汇。

虽然这些改进的 SPM 模型在性能上有所提升,但这些方法基本沿用了 SIFT 算子与 SPM 模型结合的框架,主要的改进在于码本学习的方法和空间金字塔表示方法的扩展方面。我们知道,SIFT 特征是利用了局部梯度信息进行统计,在 SPM 模型中码本学习是将 SIFT 特征在大量样本学习的基础上量化编码为词汇表示,实质上还是图像块局部梯度信息的规律表示,如果可以将 SIFT 特征统计的梯度信息进行合理编码,直接表示为词汇,则可以不进行码本学习,这将简化方法步骤,大大节省算法时间和空间成本。在图像分析领域,利用二进制编码的方法描述特征已有所应用,Mazer 等[14]对高光谱数据提出了光谱的二值编码用以描述光谱变化,被认为是一种快速有效的高光谱影像分类方法,而 Ojala 等[15]提出的 LBP(Local Binary Pattern)算子是一种能够很好地描述图像纹理特征的算子,由于它计算简单并且效果也比较好,被广泛地用于图像分类和识别。

本文通过分析 SIFT 特征描述的特点,提出了一种基于梯度方向二进制编码的特征表示,不需要由码本学习进行特征量化的步骤,显著提升了效率,并将该特征表示与 SPM 模型相结合,设计了一种新的 SPM 模型框架。实验说明了本文方法在耗时和分类性能上均超越了基于 SIFT 算子的 SPM 方法。图 1 示出了 SPM 模型和本文方法的流程框图对照。

1 SPM 模型

为了在一个特征空间中寻找两组向量的近似对应关系,Lazebnik 等[3]提出了空间金字塔模型。首先在整幅图像上提取局部特征,对局部特征进行量化得到词汇表示,再在二维图像空间根据不同尺度建立金字塔,将子图像区域的局部特征直方图加权组合为一个向量,在越精细的尺度上给予更大的权值。若金字塔的层数为 L,则在第 l 层,图像将被分为 $2^l \times 2^l$ 个同样规格尺寸的网格。

对于两幅图像 I_1 和 I_2,空间金字塔匹配核定义为:

$$K(I_1, I_2) = \sum_{l=0}^{L} \sum_{j=1}^{J_l} \omega_{l,j} K_{l,j}(I_1, I_2) \qquad (1)$$

其中,J_l 为第 l 层的网格总数,为第 l 层第 j 个网格的权值,权值设置如下:

$$\omega_{l,j}=\begin{cases}\dfrac{1}{2^{L}}, & \text{if } l=0\\ \dfrac{1}{2^{L-l+1}}, & \text{if } l>0\end{cases} \tag{2}$$

$K_{l,j}(I_1,I_2)$为根据直方图交叉核计算出的第 l 层匹配值,其定义如下:

$$K_{l,j}(I_1,I_2)=\sum_{m=1}^{M}\min(H_{l,j}(I_1),H_{l,j}(I_2)) \tag{3}$$

$H_{l,j}(I_1)$为图像 I_1 中第 l 层第 j 个网格的特征直方图,通常应用中 L 可以设为 2 或 3。

2　本文算法

如图 1(a)所示,在 SPM 模型中,首先选取一组图像作为训练样本,对每幅图像进行稠密采样划分为图像块,再使用 SIFT 特征作为局部特征描述,提取所有图像块的 SIFT 特征,然后对 SIFT 特征进行聚类学习形成码本词汇,而测试样本提取的 SIFT 特征将根据该码本量化后表示为空间金字塔特征进行分类。其中的特征提取步骤,首先使用了对局部特征进行精细描述的 SIFT 描述子,而后又对 SIFT 描述子进行简单聚类,抛弃了其中的细节信息量化为词汇。在该过程中,训练样本中包含的噪声和码本词汇分界线附近的量化会引入误差,这些都会对码本词汇自身的准确性和描述力构成影响。

若能实现一种有效的特征编码描述子,不需要进行聚类形成词汇和特征量化的步骤,这将大大提升算法效率并在一定程度上降低码本的不准确造成的影响。本文提出了梯度方向二进制模式的特征编码描述子。

2.1　梯度方向二进制模式(Binary Pattern of Oriented Gradients, BPOG)

首先,在尺度空间 L 上计算每个像素点的梯度模值和方向,

$$m(x,y)=\{[L(x+1,y)-L(x-1,y)]^2+[L(x,y+1)-L(x,y-1)]^2\}^{1/2} \tag{4}$$

$$\theta(x,y)=\arctan\frac{L(x,y+1)-L(x,y-1)}{L(x+1,y)-L(x-1,y)} \tag{5}$$

上式中,$m(x,y)$,$\theta(x,y)$分别为像素(x, y)处梯度的模值和方向,这里,L 为图像尺度空间。

接下来,我们以 8×8 的窗口为例来示意算法过程。图 2 左图为稠密采样的一个图像块,每个小格代表图像块的一个像素,箭头方向代表该像素的梯度方向,箭头长度代表梯度模值,图中圆形代表高斯加权的范围,越靠近关键点的像素梯度方向信息贡献越大。然后将图像块划分为 2×2 共 4 个子块,对于 8×8 窗口每个子块包含 4×4 个像素。在每个子块上计算 8 个方向的梯度方向直方图,计算每个梯度方向的高斯加权累加值,得到每个子块的梯度主方向,如图 2 中图所示,空心柱体为梯度主方向。接着对每个子块的梯度主方向进行二进制编码,为控制特征维度,将梯度方向共划分为 4 个区间,定义如下:梯度主方向落入 0~90°区间编码为“00”,落入 90°~180°区间编码为“01”,落入 180°~270°区间编码为“10”,落入 270°~360°区间编码为“11”,例如图 2 示例中,位于右下角子块的梯度主方向落在了 180°~270°区间,根据定义其二进制编码为“10”。

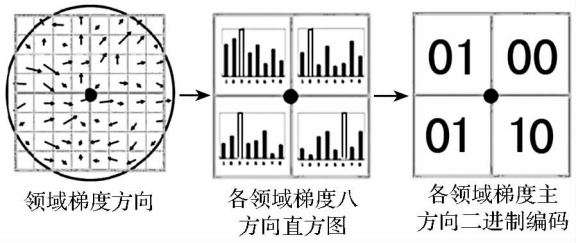


图 2　由关键点邻域梯度信息生成特征向量

Fig. 2　The process of feature generation

在得到每个小块的二进制编码后,我们需要将其表示为词汇值。如图 3 所示,将 4 个小块的二进制值逐行顺序连接为一个 8 位的二进制串值,再将其转换为十进制值,即完成了局部特征编码和词汇表示的过程。根据该编码方式,获得的字典长度为 256。这种对局部特征直接进行编码的思想避免了先学习训练形成词汇再量化的烦琐步骤。

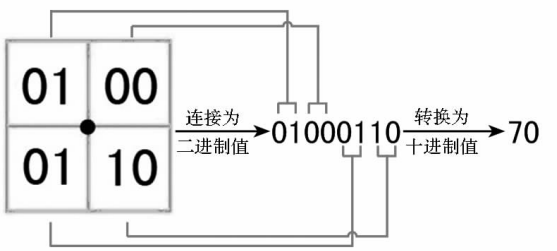


图 3　二进制编码后的词汇表示

Fig. 3　The vocabulary value of binary coding

实际计算过程中,为了增强局部特征的稳健性,在对图像进行稠密采样时,建议取 16×16 的窗口,这样划分的 4 个小块,每个小块包含 8×8 个像素。

2.2　将 BPOG 嵌入 SPM 模型

SPM 模型的基本思想是在图像空间提取局部特征后,形成视觉词汇描述的二维特征空间,再对特征空间进行一系列逐渐变细的网格划分,形

成金字塔式分割,并将每级划分下的匹配数目加权求和来得到特征集合间的相似度。本文提出的方法是将二进制编码后的词汇特征表示嵌入到 SPM 模型。

如图 4 所示,对二进制特征编码后得到的视觉词汇分别统计其在每个金字塔级别下所有网格单元内的直方图 $H_j^l(j=1,\cdots,2^{2l};\ l=0,\cdots,L)$,将所得到的各级各网格的词汇直方图加权处理并线性连接,代入到式(1)中即得到了图像的空间金字塔表示。最后将得到的空间金字塔表示代入式(3)进行匹配即形成空间金字塔匹配核。

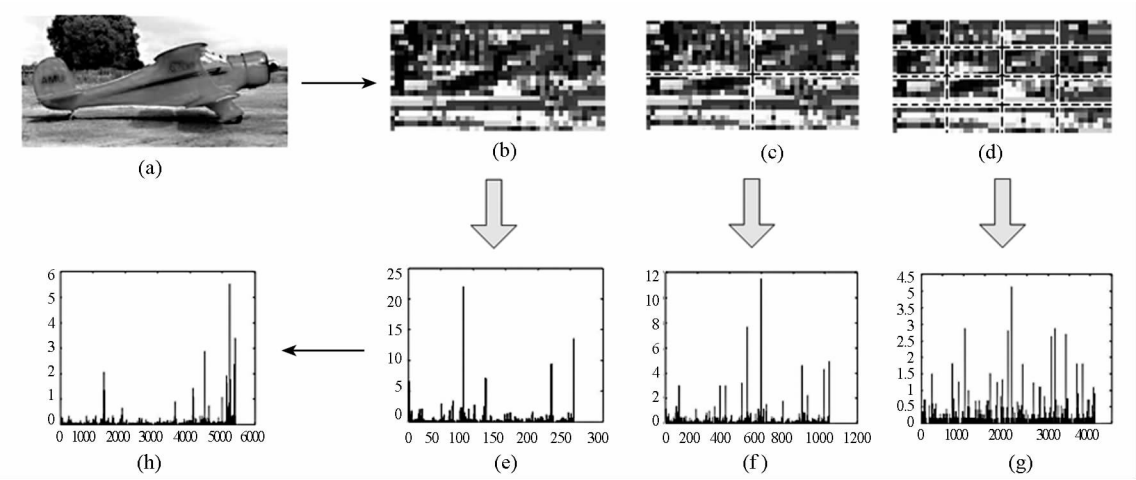


图 4 BPOG 嵌入 SPM 模型示例

(a)原始图像;(b)(c)(d)分别为金字塔层级 $l=0,1,2$ 时特征编码图上的网格划分示意;(e)(f)(g)分别对应于(b)(c)(d)的直方图表示;(h)各金字塔层级加权组合的最终直方图表示

Fig. 4 BPOG descriptor extraction in spatial pyramid model.

(a) Original image; (b)(c)(d) The grids on the BPOG code image for each level $l=0, 1, 2$, respectively; (e)(f)(g) Histogram representations corresponding to (b)(c)(d), respectively; (h) The BPOG final vector is a weighted concatenation of histograms for all levels.

本文方法视觉词汇的数目为 256,在金字塔级别为 0,1,2 时,其维数分别为 256,1024,4096,加权组合后的维数为 5376。

3 实验结果与分析

本节将介绍本文算法在常用的目标识别图像库 Caltech – 101[12] 上的分类结果。Caltech – 101 图像库共包含动物、交通工具、花等 101 个类别的图像,颜色和形状变化非常丰富。图 5 给出了该图像库中的一些样图。实验在 PC 机上进行,CPU 主频 2.4G,内存 2G,运行环境为 Windows XP 操作系统,Matlab 2009a。

在对每幅图像提取局部特征时,本文采用和基于 SIFT 特征的 SPM 方法相同的程序进行,即采用稠密采样的方式,对每幅图像在水平和垂直方向均间隔 8 个像素,提取 16 × 16 像素的图像块。为了在同等条件下对比方法,我们在方法测试中取空间金字塔层级 $L=2$,设置词汇码本长度 $M=256$。按照各文献通行的实验设置,我们在每次实验中,对每一类别随机抽取 30 幅图像加入到训练集,而将剩余图像加入到测试集。在训练和测试 SVM 分类器时,则采用了 1 对多(one-vs-all)的方式,所使用

图 5 Caltech – 101 图像库样图示例

Fig. 5 Several example images of Caltech – 101 dataset

的核函数为直方图交叉核(histogram intersection kernel)[13]。表 1 示出了本文算法与基于 SIFT 特征的 SPM 算法的结果比较,实验结果中识别率的均值和标准差为重复 10 次实验统计得到。

表 1 在 Caltech – 101 图像库上的平均准确率对比

Tab. 1 The comparison of average accuracy on Caltech – 101 dataset

算法	码本长度 M	平均准确率(%)
SIFT – SPM[3]	256	64.0615 ± 0.63
本文算法	256	65.2382 ± 0.39

特征提取的各阶段用时对比见表 2。在文献

[3]中,生成码本词汇过程通常随机抽取 50 幅图像,对这些图像的所有图像块的 SIFT 特征聚类得到,由于受计算机内存限制,在生成码本用时的实验中随机抽取了 20 幅图像共约 5 万组 SIFT 特征用来聚类计算码本。在特征描述阶段的对比中,选取了 50 幅长宽尺寸均为 300 ~ 500 像素大小的图像来测试用时,对于文献[3]方法,特征描述可以分为两步,首先提取稠密 SIFT 特征,再根据码本对 SIFT 特征量化为词汇表示,表 2 给出了这两步的各自用时,其中特征描述用时结果为一幅图像的平均用时,实验结果均为重复 10 次试验得到的平均结果。

表 2　特征提取各阶段用时对比(单位:s)

Tab. 2　The comparison of time consumption (second) on feature extraction

算法	生成码本	特征描述 (特征提取 + 量化词汇)
SIFT - SPM[3]	230	2.1524 (2.0594 + 0.0930)
本文算法	0	1.6337

由以上结果可以看出,本文算法对图像局部梯度特征进行了有效编码,在同样码本长度条件下,和基于 SIFT 特征的 SPM 算法相比,性能获得较明显提升,所获的性能提升超过 1.2%;相较于基于 SIFT 特征的 SPM 算法简化了算法步骤,在算法耗时方面也有显著提升,由于本文算法不再需要先训练生成码本,因而在生成码本阶段耗时为 0,在特征描述阶段本文算法在时间效率上提升了约 30%。

4　结论

构建有效的图像描述是计算机视觉领域的一项重要任务。在 SPM 模型中,局部特征提取后的码本学习和特征编码步骤是其中关键的环节,本文对其词汇形成过程进行分析,设计了一种优化的处理流程,将词汇的形成过程与特征提取相融合,提出了一种新的局部特征表述——梯度方向二进制模式,应用该特征描述简化了方法流程,节约了计算量。在标准的目标分类测试集上的实验结果表明,本文提出的算法由于合理利用了梯度特征信息,优化了处理流程,与基于 SIFT 特征描述的 SPM 模型方法相比,提升了算法耗时和分类的性能。在下一步工作中,我们计划对特征编码方法再做进一步研究与改进,提升特征表示的稳健性,并尝试与更多的分类模型相融合。

参考文献(References)

[1] Hofmann T. Unsupervised learning by probabilistic latent semantic analysis[J]. Machine Learning, 2001, 42(1 - 2): 177 - 196.

[2] Fergus R, Perona P, Zisserman A. A sparse object category model for efficient learning and exhaustive recognition[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005, (1):380 - 387.

[3] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories [C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006,(2): 2169 - 2178.

[4] Van Gemert J C, Veenman C J, Smeulders A W M, et al. Visual word ambiguity [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32 (7): 1271 - 1283.

[5] Yang J C, Yu K, Gong Y H, et al. Linear spatial pyramid matching using sparse coding for image classification [C]// IEEE Conference on Computer Vision and Pattern Recognition, 2009, 1794 - 1801.

[6] Boureau Y L, Bach F, LeCun Y, et al. Learning mid-level features for recognition [C]//Proceedings of Conference on Computer Vision and Pattern Recognition, 2010, 2559 - 2566.

[7] Yang J C, Yu K, Huang T. Efficient highly over-complete sparse coding using a mixture model [C]//Proceedings of European Conference on Computer Vision, 2010:113 - 126.

[8] Zhou X, Yu K, Zhang T, et al. Image classification using super-vector coding of local image descriptors [C]// Proceedings of European Conference on Computer Vision, 2010:141 - 154.

[9] Sanchez J, Perronnin F, de Campos T. Modeling the spatial layout of images beyond spatial pyramids [J]. Pattern Recognition Letters, 2012, 33(16):2216 - 2223.

[10] Jia Y Q, Huang C, Darrell T. Beyond spatial pyramids: receptive field learning for pooled image features[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2012:3370 - 3377.

[11] Yan S Y, Xu X X, Xu D, et al. Beyond spatial pyramids: a new feature extraction framework with dense spatial sampling for image classification [C]//Proceedings of European Conference on Computer Vision, 2012:473 - 487.

[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]//IEEE Conference on Computer Vision and Pattern Recognition Workshop, 2004:178.

[13] Barla A, Odone F, Verri A. Histogram intersection kernel for image classification [C]//Proceedings of International Conference on Image Processing, 2003(2):513 - 516.

[14] Mazer A S, Martin M, Lee M, et al. Image processing software for imaging spectrometry data analysis[J]. Remote Sensing of Environment, 1988, 24(1): 201 - 210.

[15] Ojala T, Pietikainen M, Maenpaa T. Multi-resolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7): 971 - 987.