

多特征融合的场景图像分类算法^①

史 静, 朱 虹

(西安理工大学 自动化与信息工程学院, 西安 710048)

摘 要: 提出了一种 Gabor-LBP 频域纹理特征与词包模型语义特征相结合的场景图像分类算法. 利用 Gabor 变换得到的频域信息, 及对应的 LBP 特征, 与视觉词包模型 (BOW) 提取的语义特征自适应相融合, 实现分类. 为了验证本文算法, 利用两个标准图像测试库进行比较测试, 实验结果表明, 本文算法在改善图像纹理表达上具有明显优势, 特别是对于图像的光照、旋转、尺度都具有很好的鲁棒性.

关键词: 场景分类; Gabor-LBP; 词包模型; SVM

引用格式: 史静, 朱虹. 多特征融合的场景图像分类算法. 计算机系统应用, 2018, 27(5): 171-175. <http://www.c-s-a.org.cn/1003-3254/6343.html>

Scene Image Classification Algorithm of Fusing Multi-Feature

SHI Jing, ZHU Hong

(School of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048, China)

Abstract: In this study, a scene image classification algorithm is proposed which combines Gabor-LBP frequency domain texture features and lexical model semantic features. The frequency domain information which obtained by Gabor transform, the corresponding LBP feature, and semantic features which extracted by the visual Bag Of Words (BOW) package model are fused to realize the classification. In order to verify the algorithm, we use two standard image test datasets to compare and test. The experimental results show that the proposed algorithm has obvious advantages in improving image texture expression, especially for image illumination, rotation, and scale.

Key words: scene classification; Gabor-LBP; bag of words; SVM

随着互联网多媒体等技术的发展, 必然带来海量的图像数据, 传统获取信息的方法已经无法满足人类的需求了, 为了对图像带来的大量数据信息进行有效的分析和处理, 需要根据图像的不同内容提前对其进行分类.

场景图像分类在机器视觉、人工智能、多媒体技术等多个领域受到广泛的关注, 比如让机器人可以像人类一样自己识别出街道、厨房、卧室等不同的场景, 更加智能的为人类服务. 此外, 场景图像分类还在图像检索^[1]、目标识别^[2]、旅游导航、数字照片自动分类、视频分类监督等领域中得到了应用. 但是由于场景图像的存在各种问题, 使得场景图像分类面临巨大

的挑战: (1) 图像纹理布局很相似这就导致图像类间比较相似; (2) 类内由于拍摄角度、近景还是远景、光照等因素的影响导致类内场景差异较大, 这些都在一定程度上这些都增加了场景图像分类的难度.

近年来, 针对场景图像分类问题已经提出了各种各样的算法, 最终目的就是, 在混合有多种类别的众多场景图像中, 提取出属于同一类场景图像之间的相同或相似特征, 如 SIFT^[3]特征、GIST^[4]特征、LBP^[5]特征等, 并用提取出的特征对这些场景图像进行尽可能准确地分类, 所以图像特征的提取是提高分类准确性中至关重要的一步.

目前的图像场景分类在一定程度上克服了人工分

^① 基金项目: 国家自然科学基金 (61502385); 西安市科技计划项目 (CX1509(13)); 西安理工大学教学研究重点项目 (xjy1775)

收稿时间: 2017-09-04; 修改时间: 2017-09-20; 采用时间: 2017-09-25; csa 在线出版时间: 2018-04-23

费时费力的弊端,但大都是建立在底层特征提取的层面上,这与人们对图像进行理解的方式有所不同.如何将底层特征进行映射成为图像高层语义特征就成为了研究的重点,本文利用 Gabor 变换^[6]在表示方向和频率时与人类的视觉系统非常相似这一点,在此基础上,为底层特征进行语义特征映射提供桥梁,将 Gabor 频域信息中提取的 LBP 特征与视觉词包模型自适应相结合构成融合特征,进行分类判别.

1 Gabor-LBP 频域纹理特征

图像在频域方面蕴含着丰富的信息,这里利用 Gabor 变换对图像进行处理. Gabor 变换其频率和方向与人类通过视觉系统看物体时的频率和方向很相似,可以很好地表现场景图像各个方向和各个尺度的更加全面的频域信息,在 Gabor 变换之后,提取 LBP 特征会得到更加丰富的纹理信息,有利于快速准确地对场景图像进行分类识别.根据经验值,选取的都是 5 个尺度和 8 个方向,一共可以组合成 40 个 Gabor 小波函数.

1.1 Gabor 变换与 LBP 结合

对图像进行 Gabor 变换就是让输入图像和 Gabor 小波核函数进行卷积运算.假设输入图像用 $I(x,y)$ 表示,那么图像 I 与 Gabor 小波核函数 $\varphi_{u,v}$ 的卷积定义可以表示为:

$$O_{u,v}(x,y) = I(x,y) * \varphi_{u,v}(x,y) \quad (1)$$

上式中的*称为卷积因子,代表的就是在方向和尺度得到的 40 幅卷积图像.

对于一幅图来说,有了 40 幅不同尺度和方向的图像,就可分别求取 LBP 特征,便可以得到更加丰富的纹理信息,将 40 幅图像分别得到的 LBP 直方图向量联合起来,就可以得到整幅图的 LBP 特征向量.图 2 是图 1 的灰度图经过 Gabor 变换后得到部分图像及提取的 LBP 特征后的部分纹理图.



图 1 场景图像 insidacity 的原图

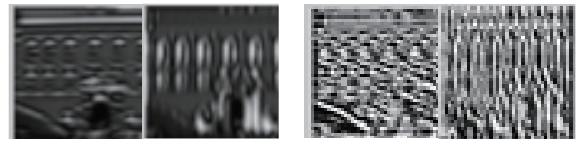


图 2 Gabor 变换图像及提取的 LBP 特征图

1.2 LBP 特征降维

上述的方法最终得到的 LBP 特征维数已经变为原来的 40 倍,相当于原来是只要求一幅图的 LBP 直方图,现在是对 40 幅图求取的 LBP 直方图才能代表一幅图.比如本文中用到的 8 类室外场景图像均为大小 256×256 的灰度图,进行 Gabor 小波变换之后求取 LBP 直方图向量,维数变到了 $256 \times 40 = 10\,240$ 维.维数一旦变得很大,对计算速度和训练分类都会有很大的影响,所以必须进行一定程度上的降维操作.

我们通过对圆形 LBP 特有的性质均匀模式研究,均匀模式不仅可以达到降维的效果,而且还不会损失原图的大量信息.使用均匀模式降维的过程中,可以将每一幅图原本提取的 256 个直方图向量映射到 59 维,那么对于一幅图像来说原来的 10 240 维就会降低到 $59 \times 40 = 2360$ 维.虽然目前的 2360 维相对于原来的 256 维来说仍然很高,但是已经在 10 240 维的基础上降低了很多,所以已经足够接下来进行分类训练.

2 词包模型的语义特征

近年来, SIFT 特征已被广泛应用于纹理特征提取方面, BOW 模型^[7]的提出将 SIFT 的良好性能用于图像分类. BOW 模型最特别的地方是把图像当作是“文档”,图像的底层 SIFT 特征点被视为视觉词汇.相关研究表明,图像的整体统计信息对于语义场景的建模非常有用,而且不用检测图像中的具体目标物.

2.1 视觉词典的构造

视觉词典^[8-10]的构建是使用所有的视觉特征集来形成可以描述这些视觉词汇的码本.视觉词典的产生是词包模型的关键,通常使用聚类方法来构建视觉词典,这里借助 K-means 聚类算法,聚类可以将具有最大相似度的特征聚为一类,聚类中心定义为“字典原子”,所有这些视觉字典原子构成一个视觉词典.

具体过程为:从训练样本中随机选出一部分图像,对其 SIFT 特征描述子进行聚类,聚类中心作为字典原子,聚类中心的集合组成视觉词典.将其中的视觉词汇

映射到与其距离最近的词典原子,其中使用视觉词汇表示图像中的局部视觉属性,统计图像中每个视觉词汇的出现频率,从而形成了基于视觉词汇表的直方图,就可以得到场景图像中的视觉词包描述.使用视觉词包模型的场景分类方法,将图像中具有类似属性的局部视觉特征转换为视觉词汇,大大降低了图像特征表达的维数,在场景图像中形成了简单有效的中层语义描述,且适应能力强,泛化能力强.

2.2 空间金字塔匹配 (SPM) 的特征描述

金字塔匹配的主要思想就是将特征空间进行网格划分,该划分是进行一系列逐渐变细,然后形成金字塔分割,并且对每个层次下的匹配数进行加权求和,以获得特征集之间的相似度.

l_0, l_1, l_2 是三个不同的层,则这几个层则被赋予不同的权值 $1/8, 1/4, 1/2$. 用 H_X^l, H_Y^l 来表示 X, Y 在 l 级下的特征直方图, $H_X^l(i), H_Y^l(i)$ 即为 X, Y 中属于第 i 网格单元的特征点,并且交叉核来描述在级别 l 下的匹配点数用直方图表示,如下式所示,将 $I(H_X^l, H_Y^l)$ 简记为 I^l .

$$I(H_X^l, H_Y^l) = \sum_{i=1}^{c(l)} \min(H_X^l(i), H_Y^l(i)) \quad (2)$$

存在于 l 层结构中的匹配点也存在于其更精细的 $l+1$ 层结构中,从而 $I^l - I^{l+1}, l=0, \dots, L-1$ 可以表示新出现在级别 l 上的匹配点.将特征空间各级划分进行总和后点形成金字塔的核函数:

$$k^L(X, Y) = I^L + \sum_{l=0}^{L-1} \frac{1}{2^{L-l}} (I^l - I^{l+1}) = \frac{1}{2^L} I^0 + \sum_{l=1}^L \frac{1}{2^{L-l+1}} I^l \quad (3)$$

对于图像来说,该模型可以有效的描述图像特征信息和空间位置信息.通过图像的多层次分割,同时结合了图像的多分辨率表示,可以在多个尺度空间中描述图像的语义特征.

3 特征融合

为了进一步挖掘图像的信息,既能提取丰富的纹理信息又能对图像各种变化有很好鲁棒性的特征,将 Gabor-LBP 和 SPM+BOW 进行融合.这里将 Gabor-LBP、SPM+BOW 分别记为 f_1, f_2 , 对这两个特征进行加权拼接,拼接后的特征称为融合特征用 F 来表示,图像的融合特征可以表示为:

$$F = [w_1 \times f_1, w_2 \times f_2] \quad (4)$$

式中 F 表示两个特征 f_1 和 f_2 的加权拼接,即为融合特征, w_1 和 w_2 分别表示 f_1 和 f_2 的权值.这里权值分配通过单个特征的识别率进行设置, f_1, f_2 的识别率分别为 A_1, A_2 :

$$w_1 = \frac{A_1}{A_1 + A_2}, w_2 = \frac{A_2}{A_1 + A_2} \quad (5)$$

为了避免不同的特征的数量级差别较大在特征融合时数量级较大的淹没数量级较小的特征,这里我们对 Gabor-LBP, SPM+BOW 两种特征进行归一化,然后按照上述中所说给每个特征赋予相应的权值进行加权拼接,拼接后的特征作为图像的特征进行分类识别.分类器我们使用一对一 SVM 分类器,由于一对多在设计时是将一类和其余类训练一个分类器这样常常会出现数据失衡的问题造成最后的识别率并不可靠.训练完分类器后,对于测试图像也经过相同的过程提取图像的融合特征,归一化后加权拼接,最后输入到分类器进行分类决策.

4 实验结果分析

4.1 实验数据

为了验证本章提出的算法对于图像场景分类的性能优越性.本文采用了由 Oliva 等人提供的 8 类室外场景数据集^[11](OT 数据集); Li Fei-Fei 等人提供的 8 类运动场景^[12](SE 数据集).所有数据的准确率均为 5 次试验的平均结果.

1) OT 数据集

该数据集总共包括 2688 张图片,包含 coast、forest、mountain 等,共 8 类自然场景,且每幅图像都是大小为 256×256 的灰度图,部分图像如图 3 所示,其中每类 100 张训练,其余测试.



图 3 OT 数据集

2) SE 数据集

该数据集总共包括 1579 张图片,其中包含的运动场景有: badminton、bocce、croquet 等 8 类运动场景,

每类图像的数目从 137 张到 250 张不等. 部分图像如图 4 所示, 其中每类 70 张训练, 其余 60 张测试.



图 4 SE 数据集

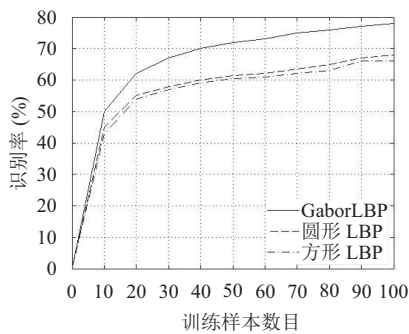
4.2 对比实验

为了验证 Gabor-LBP 特征的有效性, 这里对方形 LBP, 圆形 LBP, Gabor-LBP, 依次进行测试. 各特征分类结果如表 1 所示.

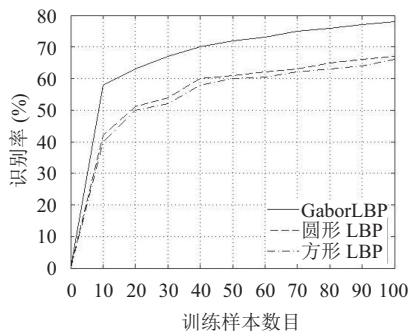
表 1 各数据集改进 LBP 与基本 LBP 方法的正确率比较 (%)

	方形LBP	圆形	Gabor-LBP
OT	66.10	67.90	79.01
SE	59.56	61.30	74.87

为了探究不同训练数目下各个特征的识别率, 我们对 OT 和 SE 数据集不同训练样本数目进行测试, 我们以训练样本从 10 到 100 逐渐增加实验结果如图 5 所示.



(a) OT 数据集



(b) SE 数据集

图 5 各数据集训练样本对识别率影响的曲线图

从图 5 可以看出随着训练样本的逐渐增加识别率逐渐提高, 但 Gabor-LBP 的识别率始终高于其他 LBP 特征的识别率.

为了验证融合特征与单个特征分类识别率的比较实验, 对两个数据集进行比较测试, 结果如表 2 所示.

表 2 各数据集整体识别率结果对比表 (单位: %)

	Gabor-LBP	SPM+BOW	融合特征
OT	83.1	83.6	86.4
SE	74.2	76.5	83.7

最后, 将本文方法与其他参考文献方法的进行比较, 结果如表 3 所示.

表 3 不同方法结果对比表 (单位: %)

方法	OT数据集	SE数据集
文献[6]	86.2	78.3
文献[3]	79.1	67.1
文献[13]	83.4	—
文献[9]	79.9	67.4
本文方法	86.4	83.7

由表 3 可以看出本文的图像融合的方法在场景图像分类方面有一定的优势, 由于特征融合提供了更为丰富的图像信息使得图像更具有区分性, 从而提高了图像的识别率.

5 结论

本文提出了新的场景图像分类算法, 将图像的 Gabor-LBP 特征与视觉词包模型自适应相结合, 进一步挖掘图像的信息, 可以获得对场景图像的多个尺度多个方向的丰富纹理特征, 以及在图像尺度, 旋转, 光照有很好的鲁棒性特征. 场景图像的标准数据集的测试结果验证了本文方法的有效性.

参考文献

- Xue ZY, Rahman MM, Antani S, et al. Modality classification for searching figures in biomedical literature. Proceedings of the 29th International Symposium on Computer-Based Medical Systems (CBMS). Dublin, Ireland. 2016. 152-157.
- Shao L, Liu L, Li XL. Feature learning for image classification via multiobjective genetic programming. IEEE Transactions on Neural Networks and Learning Systems, 2014, 25(7): 1359-1371. [doi: 10.1109/TNNLS.2013.2293418]
- Lowe DG. Distinctive image features from scale-invariant

- keypoints. *International Journal of Computer Vision*, 2004, 60(2): 91–110. [doi: [10.1023/B:VISI.0000029664.99615.94](https://doi.org/10.1023/B:VISI.0000029664.99615.94)]
- 4 Yin JH, Li H, Jia XP. Crater detection based on GIST features. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015, 8(1): 23–29. [doi: [10.1109/JSTARS.2014.2375066](https://doi.org/10.1109/JSTARS.2014.2375066)]
- 5 Nanni L, Lumini A, Brahnam S. Survey on LBP based texture descriptors for image classification. *Expert Systems with Applications*, 2012, 39(3): 3634–3641. [doi: [10.1016/j.eswa.2011.09.054](https://doi.org/10.1016/j.eswa.2011.09.054)]
- 6 Tao DC, Li XL, Wu XD, *et al.* General tensor discriminant analysis and Gabor features for gait recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(10): 1700–1715. [doi: [10.1109/TPAMI.2007.1096](https://doi.org/10.1109/TPAMI.2007.1096)]
- 7 Kejriwal N, Kumar S, Shibata T. High performance loop closure detection using bag of word pairs. *Robotics and Autonomous Systems*, 2016, (77): 55–65. [doi: [10.1016/j.robot.2015.12.003](https://doi.org/10.1016/j.robot.2015.12.003)]
- 8 Zhou L, Zhou ZT, Hu DW. Scene classification using a multi-resolution bag-of-features model. *Pattern Recognition*, 2013, 46(1): 424–433. [doi: [10.1016/j.patcog.2012.07.017](https://doi.org/10.1016/j.patcog.2012.07.017)]
- 9 Chu WT, Chen CH, Hsu HN. Color CENTRIST: Embedding color information in scene categorization. *Journal of Visual Communication and Image Representation*, 2014, 25(5): 840–854. [doi: [10.1016/j.jvcir.2014.01.013](https://doi.org/10.1016/j.jvcir.2014.01.013)]
- 10 Jeong DJ, Yoo HJ, Cho NI. Consumer video summarization based on image quality and representativeness measure. *Proceedings of 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*. Orlando, FL, USA. 2015. 572–576.
- 11 Oliva A, Torralba A. Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision*, 2001, 42(3): 145–175. [doi: [10.1023/A:1011139631724](https://doi.org/10.1023/A:1011139631724)]
- 12 Li LJ, Li FF. What, where and who? Classifying events by scene and object recognition. *Proceedings of the 2007 IEEE 11th International Conference on Computer Vision*. Rio de Janeiro, Brazil. 2007. 1–8.
- 13 Wu JX, Rehg JM. CENTRIST: A visual descriptor for scene categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(8): 1489–1501. [doi: [10.1109/TPAMI.2010.224](https://doi.org/10.1109/TPAMI.2010.224)]