

文章编号:1001-9081(2009)09-2393-02

图学习的区域图像标注方法

虎晓红^{1,2}, 钱旭², 王珂³

(1. 河南农业大学 信息与管理科学学院, 郑州 450002; 2. 中国矿业大学 机电与信息工程学院, 北京 100083;
3. 首都医科大学 生物医学工程学院, 北京 100069)
(wd9702@163.com)

摘要:近年来, 图像标注技术得到广泛关注。提出一种图学习的自动图像标注方法, 将图像标注作为多示例学习框架下的半监督学习策略, 通过给出适合图像在包空间的有效度量方式, 充分利用未标注样本挖掘图像特征的内在规律性, 将半监督学习的方法和多示例学习有效结合起来, 从而获得更准确的标注结果。实验结果表明, 提出的标注方法可行, 同时标注结果与传统的标注方法相比得到了明显提高。

关键词:多示例学习; 半监督学习; 自动图像标注; 图学习; 区域匹配

中图分类号: TN911.73; TP391 **文献标志码:** A

Region-based image annotation of graph learning approach

HU Xiao-hong^{1,2}, QIAN Xu², WANG Ke³

(1. College of Information and Management Science, Henan Agricultural University, Zhengzhou Henan 450002, China;
2. School of Mechanical Electronic and Information Engineering, China University of Mining and Technology, Beijing 100083, China;
3. School of Biomedical Engineering, Capital Medical University, Beijing 100069, China)

Abstract: Image annotation has been an active research topic in recent years. The authors formulated image annotation as a semi-supervised learning problem under multi-instance learning framework. A novel graph-based semi-supervised learning approach to image annotation, using multiple instances, was presented, which extended the conventional semi-supervised learning to multi-instance setting by introducing the adaptive geometric relationship between two bags of instances. The experimental results show that this approach outperforms other traditional methods and is effective for image annotation.

Key words: multi-instance learning; semi-supervised learning; automatic image annotation; graph learning; region matching

0 引言

随着科技的不断发展及图像数据急剧增长, 早期手工标注方式的主观性及费时性已使其无法有效地实现对图像库的语义标注, 因此, 自动图像标注技术^[1-3]近年来得到许多关注并成为重点研究领域之一。根据图像特征所表达信息的不同, 图像自动标注可分为基于全局特征的标注和基于区域特征的标注, 相对于全局的方法, 区域特征更符合人的视觉特点。然而, 在基于区域特征的自动标注中, 由于标注对应的是整幅图像而非每个区域, 使得图像标注更为复杂。同时, 图像的手工标注样本费时费力, 据此, 在自动图像标注中, 标注样本的获取代价较高, 使得有限的已标注图像成为自动标注中的另一主要问题。文献[4]运用图学习的方法采用分割后的区域特征^[5], 提出了 NSC 模式生成自适应图的自动标注方法。本文采用固定大小网格的图像区域特征, 在图像匹配阶段, 通过融合各子块匹配值, 给出适合图像在包空间的有效度量方式来挖掘图像特征的内在规律, 同时运用已标注和未标注图像^[6]在包空间中对图像库中的每幅图像进行自动标注以提高标注性能。

1 图像表示及匹配策略

按图像分割的粒度不同, 图像区域可分为分割后区域和

固定大小的网格。在图像标注中, 分割后的区域由于受图像分割技术的影响, 易形成弱语义区域, 影响标注效果; MBRM^[7]中将基于固定大小的网格与基于分割的标注性能进行了比较, 实验结果表明, 基于固定网格的图像划分的标注性能优于基于图像分割大小的标注性能, 据此, 本文采用基于图像网格的区域特征。设图像库由 $X = \{x_1, x_2, \dots, x_n\} \subset \mathbf{R}^m$ 构成, n 为图像库中图像的数目, m 为特征空间的维数, 将图像库中的每幅图像分为 3×3 大小的网格, 分别对每个图像网格进行颜色和纹理视觉特征提取, 图像包空间的每个示例的视觉特征为 HSV 空间下 $8 \times 2 \times 2$ 量化级数的颜色直方图和 $0^\circ, 45^\circ, 90^\circ, 135^\circ$ 以及无方向 5D 的边缘直方图特征。

区域匹配技术对图像标注性能的影响至关重要, 本文中, 据 FUZZYCLUB^[8]的思想, 定义图像 $x_j (1 \leq j \leq n)$ 的某子块 $b_m (1 \leq m \leq 9)$ 和另一幅图像 $x_i (1 \leq i \leq n)$ 的距离为:

$$d(b_m, x_i) = \min \| b_m - a_p \| \quad (1)$$

即一图像子块到另一图像的距离为该子块到另一图像子块距离的最小值, 其中, $a_p (1 \leq p \leq 9)$ 为 x_i 中的子块; 图像包的距离为一个包中的每个示例和另一包中的最近示例的距离融合值:

$$d(x_i, x_j) = \frac{\sum_{m=1}^9 \text{dist}(b_m, x_i) + \sum_{p=1}^9 \text{dist}(a_p, x_j)}{|x_i| + |x_j|} \quad (2)$$

收稿日期: 2009-03-23; 修回日期: 2009-05-08。 基金项目: 教育部科研重点项目(107021)。

作者简介: 虎晓红(1978-), 女, 河南信阳人, 讲师, 博士研究生, 主要研究方向: 信息融合、机器学习、图像检索; 钱旭(1962-), 男, 江苏徐州人, 教授, 博士生导师, 主要研究方向: 信息融合、知识工程、计算机支持的协同工作; 王珂(1978-), 女, 河南信阳人, 实验师, 硕士, 主要研究方向: 图像检索。

2 基于图学习的图像标注

在图像标注中,有限的已标注图像是自动标注中的一主要问题,基于图学习的算法是一种半监督算法,在学习过程中充分利用已知类标的训练数据和未知类标的测试数据,在学习阶段使用了更多的信息。其主要思想是利用数据的总体空间分布特征和原始类标信息,使得最终得到的分类结果在数据空间总体上能够充分平滑(即相邻点的类标信息相似),同时保证尽可能地拟合训练数据。

2.1 图的建立

由图像库中的 n 幅图像为节点构建图,在创建图的过程中,采用 K 近邻(K-Nearest Neighbor, KNN)的方法构建图,计算包空间的 K 近邻点,连接 x_i 的 K 近邻点作为加权图的边。图的边权值矩阵为:

$$W_{ij} = \exp(-d(x_i, x_j)^2 / (2\sigma)^2) \quad (3)$$

其中 σ 为给定的参数,由式(2)计算 $d(x_i, x_j)$,为了防止自相似性,设 $W_{ii} = 0$ 。据 $S = D^{-1/2} W D^{-1/2}$ 规一化 W ,其中 D 为对角

阵,满足 $D_{ii} = \sum_{j=1}^n W_{ij}$ 。

2.2 标注的传递

由数据点在包空间中构造加权图后,在图学习中,利用迭代将数据点的标注通过加权图逐渐传递给邻近的点。对图像库中的图像 x_i ,设标注集为 $L = \{1, 2, \dots, C\}$,初始化 $n \times c$ 的矩阵 T ,其值为二值向量,满足:

$$T_i = \delta_c(x_i); \forall i \in \{1, 2, \dots, n\} \quad (4)$$

在标注自动传递的每一步,每个图像融合其近邻图像的标注信息,据此,在 $q+1$ 时刻,图像库的图像的排序评分为:

$$F_{q+1} = \alpha S F^q + (1 - \alpha) T \quad (5)$$

此处, α 为参数并且满足 $\alpha \in (0, 1)$,它定义了标注传递过程中,某点所获得的标注信息的来源, $\alpha S F^m$ 表示它从邻近点所获得的标注信息, $(1 - \alpha) T$ 表示它从样例数据点所获得的标注信息;排序按照式(5)做迭代直至 F 收敛,文献[9]证明了序列 $\{F^m\}$ 收敛于:

$$F^* = (1 - \alpha)(I - \alpha S)^{-1} T \quad (6)$$

3 基于图学习的区域图像检索算法

检索算法步骤如下。

步骤1 将图像库中的图像分为固定大小的网格,提取每个图像区域的颜色,纹理视觉特征,依据区域特征生成包空间。

步骤2 计算包空间的 K 近邻点,连接 x_i 的 K 近邻点作为加权图的边。

步骤3 计算包空间的权值矩阵 W ,如果 x_i, x_j 相连,则 $W_{ij} = \exp(-d(x_i, x_j)^2 / (2\sigma)^2)$,其中据式(2)计算包空间 x_i, x_j 距离 $d(x_i, x_j)$,否则 $W_{ij} = 0$,为了防止自相似,设定 $W_{ii} = 0$ 。

步骤4 据 $S = D^{-1/2} W D^{-1/2}$ 归一化 W ,其中 D 为对角阵,满足 $D_{ii} = \sum_{j=1}^n W_{ij}$ 。

步骤5 按照公式 $F_{m+1} = \alpha S F^m + (1 - \alpha) T$ 做迭代直至 F 收敛,其中 α 是给定的参数,其取值为 0 和 1 之间。

步骤6 设 F^* 为最后得到的收敛的结果,其中的元素 F_i^* 对应 x_i 的标注评分。

4 实验结果及分析

实验数据集为 corel 图像库,其中有花,熊,虎,人物,飞

机,山脉,鸟,珊瑚等 50 类图像,每类中包含 100 张图像,共计 5000 张图像作为实验数据,每幅图像与 1~5 个标注词关联,共 371 个词,出现在测试集上的标注词为 260 个,其中 4500 幅作为训练图像,500 幅作为测试图像。我们将每幅图像分为 3×3 个区域,区域特征为 32D 的颜色特征,5D 的纹理特征。实验采用标注准确率,标全率, $F1$ 作为评价标准。

图 1 为所提出的方法在 corel 图像库中标注性能图,横轴表示标注词的数量,纵轴表示标注准确率,标全率及 $F1$ 相应的百分数,表 1 为所提出的方法和 MBRM 方法在 corel 图像库中对标注词长度为 5,在 260 个标注词上的性能比较表,从表中可以看出,所提出的方法在标注性能上明显高于 MBRM 方法, $F1$ 提高了 2%,标注词(标全率 > 0)增加 13 个。表 2 为所提出的方法和 MBRM 方法在 corel 图像库中对标注词长度为 5,在 70 个常用标注词^[4]上的性能比较,从表 2 中可以看出,所提出的方法在标注性能上高于 MBRM 方法。

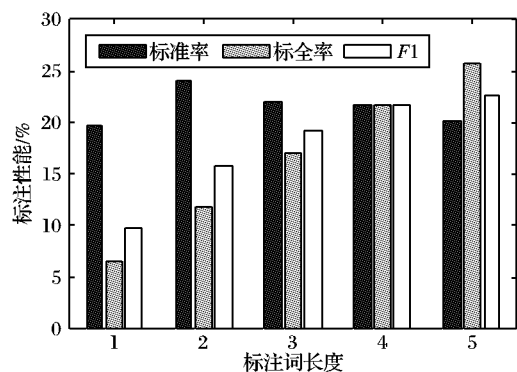


图 1 图像标注结果

表 1 图像库标注性能比较(260 标注词)

算法	标准率/%	标全率/%	F1/%	标注词 (标全率 > 0)
MBRM	18.6	23.0	20.6	118
本文方法	20.0	25.7	22.6	131

表 2 图像库标注性能比较(70 标注词)

算法	标准率/%	标全率/%	F1/%	标注词 (标全率 > 0)
MBRM	33.4	44.0	38.0	66
本文方法	35.0	44.2	39.1	67

5 结语

本文提出了一种基于图学习的区域图像标注方法。为了获得更好的标注性能,提出了一种适合图像标注的区域匹配方法,同时运用图学习的方法,整体考虑所有图像在特征空间的相关性,利用特征空间中的潜在流行分布结构,对相应图像标注传递。实验结果表明,本文的图像标注方法是可行的,同时标注性能得到了明显的提高。

参考文献:

- [1] DUYGULU P, BARNARD K, de FREITAS J F G. Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary [C]// Proceedings of 2002 European Conference on Computer Vision. Berlin: Springer-Verlag, 2002: 97-112.
- [2] JEON J, LAVRENKO V, MANMATHA R. Automatic image annotation and retrieval using cross-media relevance models [C]// Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2003: 119-126. (下转第 2397 页)

集1中选出每人2幅图像作为训练样本,进行了相同的实验,结果如表3所示,其中 $\lambda = 0.17$,2级分解。若将识别直接在对数域下进行,平均结果还将得到一定的改进。

表1 不同 λ 取值对识别率的影响 %

λ	子集2	子集3	子集4	子集5
0.01	100.00	99.17	96.43	96.84
0.02	100.00	100.00	98.57	99.47
0.04	100.00	100.00	99.29	98.42
0.05	100.00	100.00	100.00	98.42
0.08	100.00	100.00	100.00	99.47
0.09	100.00	100.00	100.00	100.00
0.10	100.00	100.00	100.00	100.00
0.18	100.00	100.00	100.00	100.00
0.19	100.00	100.00	99.29	100.00
0.30	100.00	100.00	98.57	100.00
0.40	98.33	100.00	97.14	100.00
0.45	97.50	100.00	95.71	100.00
0.50	95.83	100.00	93.57	96.32
1.00	94.17	100.00	93.57	98.42

表2 不同方法的识别率比较 %

方法	子集2	子集3	子集4	子集5
MSR	100	100.00	99.29	100.00
SQI	100	98.33	87.86	96.84
LTV	100	100.00	94.29	92.63
本文方法	100	100.00	100.00	100.00

表3 部分小波小样本平均识别率 %

小波	子集1	子集2	子集3	子集4	子集5	平均值
haar	100	100	99.17	98.57	97.37	99.02
biorl.1	100	100	99.17	98.57	97.37	99.02
biorl.3	100	100	100.00	99.29	99.47	99.75
biorl.5	100	100	100.00	98.57	100.00	99.71

图3所示为以子集1中图像为训练样本时,子集4和子集5上识别率随样本数增加的变化曲线,子集1、2、3上的识别率均为100%。数据来自 $\lambda = 0.17$ 时sym2小波3级分解。

3 结语

本文提出了一种针对光照变化条件下的多尺度光照不变量提取的人脸识别方法。该方法通过对数和小波变换的多尺度特性对因光照影响而产生的人脸图像的不同灰度尺度空间所造成的识别困难给出了有针对性的解决方法,并在Yale B人脸库上取得了较好的实验效果。该方法中仅需视光照情况确定一个参数 T ,但在其取值范围内均可以取得较理想的识

别效果,具有较好的鲁棒性。该方法易实现,对小样本问题适用,具有一定的实用价值。

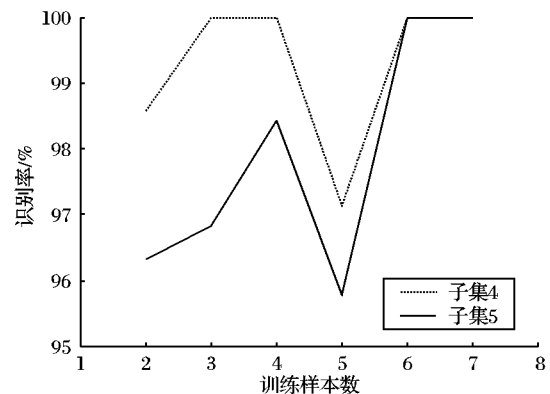


图3 识别率随样本数增加的变化曲线

参考文献:

(上接第2394页)

- [1] ZHAO W, CHELLAPPA R, ROSENFELD A, *et al.* Face recognition: A literature survey [J]. ACM Computing Surveys (CSUR), 2003, 35(4): 399-458.
- [2] ADINI Y, MOSES Y, ULLMAN S. Face recognition: The problem of compensating for changes in illumination direction [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 721-732.
- [3] CHEN T, YIN W, ZHOU X S, *et al.* Total variation models for variable lighting face recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(9): 1519-1524.
- [4] HORN B K P. Robot vision[M]. Cambridge, MA: MIT Press, 1986.
- [5] DONOHO D L. De-noising by soft-thresholding [J]. IEEE Transactions on Information Theory, 1995, 41(3): 613-627.
- [6] DONOHO D L, JOHNSTONE I M. Adapting to unknown smoothness via wavelet shrinkage [J]. Journal of the American Statistical Association, 1995, 90(432): 1200-1224.
- [7] CHANG S G, YU B, VETTERLI M. Adaptive wavelet thresholding for image denoising and compression [J]. IEEE Transactions on Image Processing, 2000, 9(9): 1532-1546.
- [8] JOBSON D J, RAHMAN Z U, WOODSELL G A. A multiscale retinex for bridging the gap between color images and the human observation of scenes [J]. IEEE Transactions on Image Processing, 1997, 6(7): 965-976.
- [9] WANG HAI-TAO, LI S Z, WANG YANG-SHENG. Face recognition under varying lighting conditions using self quotient image [C]// Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition. Washington, DC: IEEE Computer Society, 2004: 819-824.
- [7] FENG S L, MANMATHA R, LAVRENKO V. Multiple Bernoulli relevance models for image and video annotation [C]// Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC: IEEE Computer Society, 2004: 1002-1009.
- [8] ZHANG R F, ZHANG Z F. A clustering based approach to efficient image retrieval [C]// Proceedings of the 14th IEEE Conference on Tools with Artificial Intelligence. Washington, DC: IEEE Computer Society, 2002: 339-346.
- [9] ZHOU D, BOUSQUET O, GRETTON A, *et al.* Ranking on data manifolds [EB/OL]. [2009-01-05]. <http://research.microsoft.com/en-us/um/people/denzho/papers/rdm.pdf>.

- [3] JIN R, CHAI J Y, SI L. Effective automatic image annotation via a coherence language model and active learning [C]// Proceedings of 2004 International Conference on ACM Multimedia. New York: ACM Press, 2004: 892-899.
- [4] 卢汉清, 刘静. 基于图学习的自动图像标注[J]. 计算机学报, 2008, 31(9): 1629-1639.
- [5] SHI J, MALIK J. Normalized cuts and image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 2(8): 888-905.
- [6] BLUM A, CHAWLA S. Learning from labeled and unlabeled data using graph mincuts [C]// Proceedings of the 18th International Conference on Machine Learning. New York: ACM Press, 2001: 19-26.