

全局及其个性化区域特征的图像检索

段 娜 王 磊

(公安部第三研究所 上海 200000)

摘 要 交通领域个性化图像检索的关键是根据业务需求通过重点监控车辆的个性化特征在海量数据库中进行匹配,其目的是捕获与重点监控车辆相关的卡口信息。目前的图像检索算法包括基于文本的图像检索、基于内容的图像检索方法和基于语义的图像检索。针对交通领域的图像检索需求,提出了一种基于全局以及个性化感兴趣区域特征的图像检索算法。通过使用交通图像库进行检索验证,对个性化特征进行精准滤除,从而得到准确的检索结果。实验表明,此种基于全局特征结合个性化感兴趣区域特征的图像检索算法解决了 CNN 高层特征对个性化局部特征描述能力低、检索耗时等问题,并通过个性化局部特征提高了检索效果,使得检索率、平均准确率都达到 90%,呈现出较好的检索效果,计算速度快,具有较强的鲁棒性和实用性。

关键词 卷积神经网络, HOG 特征, 图像检索, 智能交通系统

中图法分类号 TP391 文献标识码 A

Image Retrieval of Global and Personalized ROI Adjustment of Features

DUAN Na WANG Lei

(The Third Research Institute of The Ministry of Public Security, Shanghai 200000, China)

Abstract The key of personalized image retrieval in the area of traffic is to capture the bayonet related information of the key monitoring vehicle through their personalized features. The current image retrieval algorithms include text based image retrieval, content-based image retrieval and semantic based image retrieval. Based on the requirements of traffic in the field of image retrieval, a new image retrieval algorithm based on global and individual interest region features was presented. To get accurate search results, we used the traffic image database to search, verify and filter the accurate personalized features. The experiment demonstrates that this method solves the problem of CNN features, such as low ability to describe personalized features and time-consuming. This method also has strong ability to describe personalized features, both the image retrieval rate and the average accuracy rate reach 90%, showing a better retrieval effect, fast calculation speed, strong robustness and practicality.

Keywords Convolutional neural network, Histograms of oriented gradients, Image retrieval, Intelligent transportation system

1 引言

近年来,随着各地智能交通事业的不断扩大,治安卡口、及准卡口爆炸式增长的图像数据给交通治安带来了丰富便利的信息,对提高城市治安管理效率、维护城市安全起了重大作用。但是如何有效组织并从海量视频图像中快速准确地进行车辆轨迹查询、交通违法查询、黑名单报警等成为图像检索领域亟待解决的问题。目前,图像检索领域有 3 种图像检索方法:1)基于文本的图像检索^[1](Test-Based Image Retrieval, TBIR);2)基于内容的图像检索^[2-4](Content-Based Image Retrieval, CBIR);3)基于语义的图像检索。基于文本的图像检索和基于内容的图像检索方法是主观耗时的;基于语义的图像检索是为了使图像尽可能逼近人对图像内容的理解,采用机器学习等技术进行检索^[5,6]。视觉信息能够客观地分析高

层语义和缩小语义鸿沟,因此,基于视觉信息的图像检索能有效地解决存在的语义问题。目前深度学习算法为当前机器学习中一个核心的分支,是图像处理中解决海量多维度特征数据效果最好的算法。深度学习主要强调了神经网络的深度^[7,8],还突出了特征学习的重要性,从大数据中学习特征,这些特性表达了丰富的内在信息,消除了不同的底层特征对检索结果带来的影响。

传统的卡口平台通过记录点位、时间段、车牌号、车辆类型、车身颜色等信息查询和监督车辆往来记录。然而,由于车牌污损、无牌车、识别错误等因素的存在,车牌识别功能有时无法提供上述基础数据,无法确认车辆身份,此时,由于缺少相关充分的数据,卡口的查询功能对于这些车辆往来很难实现有效的检测与监督。对于重点检测的车辆,常使用故意悬挂套牌、假牌或者故意不悬挂车牌的手段隐瞒车辆身份,这种

本文受公安部技术研究计划项目(2015JSYJB26),公安部技术研究计划竞争性遴选项目(2014QZX005),江苏省科技支撑计划(BE2014646),苏州市科技支撑计划(SS201413)资助。

段 娜(1985—),女,硕士,工程师,主要研究方向为视频图像智能分析等,E-mail:naduan323@163.com;王 磊(1987—),男,硕士,主要研究方向为系统架构、模式识别等。

情况反而使前端卡口系统对重点监测的车辆失去了有效的监测手段。基于上述问题,针对交通卡口图像数据,本文提出了基于全局及其个性化感兴趣区域特征的图像检索方案。

交通领域个性化图像检索的关键就是根据业务需求重点监控车辆的个性化特征,捕捉重点监控车辆相关的卡口信息。这些兴趣区域及其空间关系准确地构建了车辆的个性化文件来进行个性化滤除。视觉感知信息为用户提供了精确的个性化感兴趣信息,其不仅内容客观丰富,而且是自然高效和透明的捕捉方法。本文提出的个性化检索算法为基于视觉感知的个性化图像检索技术,其目的是通过直接感知视觉信息和客观地描述用户的语义来缩小语义鸿沟,并准确地提取视觉特征用于个性化滤除。

2 个性化检索方案

2.1 基于 CNN 特征的图像检索

本文采用文献[9]中提出的深度神经网络 AlexNet 对卡

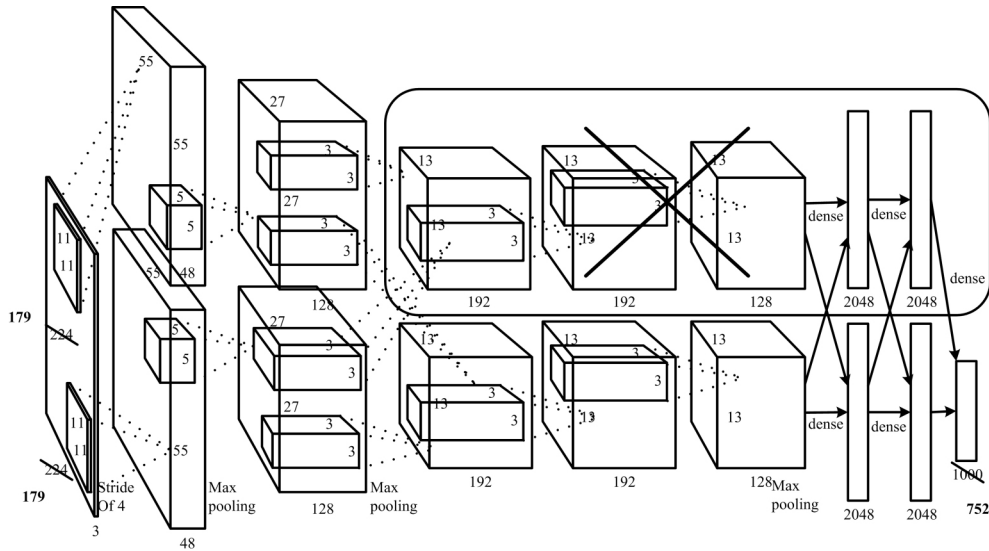


图1 缩减后的 AlexNet 结构

数据集包含训练样本 150 万张,测试样本 20 万张,共 1244 个车辆品牌标签类别。Compressed Net 网络模型在 CPU Intel i7-3770 条件下,测试数据集为 4 万张,测试结果如表 1 所列,准确率比 AlexNet 降低 1%,但是平均处理速度提升到 40ms。

表 1 识别结果

模型	准确率(%)	耗时(ms)
AlexNet	98.1	100
Compressed AlexNet	97.3	40

使用 Compressed Net 模型的 fc6 层作为全局特征用于模板图像的检索。

用式(1)计算 fc6 层特征的相似性。

$$dist(F_{obj}, F_{query}) = \sqrt{\sum_{i=0}^m (F_{obj,i} - F_{query,i})^2} \quad (1)$$

依据待检测图像与模板图像的 fc6 层特征的欧氏距离得出它们的相似度。

2.2 底层视觉特征的个性化检索

交通领域的个性化图像检索的关键是根据业务需求重点监控车辆的个性化特征,捕获重点监控车辆相关的卡口信息。底层特征能够准确地描述个性化兴趣区域,通常底层特征的

口车辆图像训练品牌细分类模型,利用车辆的品牌分类模型计算的高层特征,对待检索图像进行第一次滤除。

AlexNet 网络结构比较简单,主要由 5 个卷积层,3 个全连接层和 Soft-max 层组成,部分卷积层后紧跟着 Max-Pooling 层。该网络应用了非线性的激活函数 ReLU,重叠池化(Over-lapping Pooling)的方式在某种程度上降低了过拟合几率。

在一定的准确率损失下,通过近似和简化 AlexNet 网络加速 AlexNet 计算(见图 1)。第一个 conv 层(conv1)从 $256 \times 256 \times 3$ 的输入图像变为 $179 \times 179 \times 3$;第三层(conv3)与第四层(conv4)卷积层的个数从 384 降低到 196;第五层(conv5)卷积层的个数从 256 降低到 128;第一个 full-connected 层(fc6)为 $128 \times 6 \times 6$,本层的最终节点数为 2048;同时第二个 full-connected 层(fc7)最终节点数为 2048。这样不仅测试计算快速,而且还可以节省训练时间。

图像匹配方法是利用图像特征描述子来表示图像个性化区域和模板图像信息,通过衡量特征描述子之间的相似度来寻找个性化特征的最佳匹配位置及特征。SIFT^[10]和 SURF^[11]均给出了一种特征描述方法,其具有光照和旋转不变性,但是构造特征描述子的过程较复杂,计算量大,难以满足实时性要求。本文采用了基于空间的 HOG 特征^[12]作为特征描述子用于个性化特征筛选。

2.2.1 归一化处理

保证模板图像的信息及兴趣区域与待检索图像的一致性,提高匹配的准确率。

(1)模板图像的归一化处理如式(2)所示,根据 $ratio_{norm}$ 将图像的 size 做归一化处理,得到归一化后的图像 Obj_{norm} 。

$$ratio_{norm} = \frac{500}{width_{obj_img}} \quad (2)$$

$$Obj_{norm}(i, j) = ratio_{norm} \times Obj_{img}(i, j) \quad (3)$$

(2)根据式(2)的比率对 ROI 做归一化处理,得到模板图像的感兴趣区域 $rect_{norm}$ 。

$$rect_{norm} = ratio_{norm} + rect_{ori} \quad (4)$$

(3)目标图像的亮度归一化处理。

$$B_{norm}(i, j) = \alpha \times Obj_{rect_{norm}}(i, j) \quad (5)$$

$$\alpha = \begin{cases} \frac{B_1+10}{Avg_{bri}}, & Avg_{bri} \leq B_1 \\ \frac{B_2-10}{Avg_{bri}}, & Avg_{bri} \geq B_2 \\ 1, & \end{cases} \quad (6)$$

其中, Avg_{bri} 表示 Lab 色彩空间中 L 的平均亮度, B_1, B_2 表示亮度经验值, $Object_{norm}$ 表示归一化后的 ROI 图像。

2.2.2 提取模板的特征

HOG 特征通过提取局部区域梯度分布表示局部区域内梯度分布特征。提取模板图像 $Object_{norm}$ 的 HOG 特征, HOG 特征描述子提取过程分为 3 步:

(1) 用式(7)、式(8)计算梯度幅值和方向。

$$m(i, j) = \sqrt{G_i^2 + G_j^2} \quad (7)$$

$$\theta(i, j) = \arctan\left(\frac{G_j}{G_i}\right) \quad (8)$$

其中, G_j, G_i 分别代表水平和竖直方向的梯度。梯度方向 $\theta(i, j)$ 可以设定为 $0 \sim \pi$ 或者 $0 \sim 2\pi$ 。梯度的方向反映该像素点周围的灰度变化方向, 梯度的幅度 $m(i, j)$ 反映灰度变化的大小。

(2) 将图像划分为若干个图像块(BLOCK), 每个块划分为若干个正方形图像单元(CELL)。以 CELL 为单位统计梯度直方图, 形成一个 BIN 维的向量。

(3) 以 BLOCK 为单位进行标准化, 减低光照不均的影响。设 V 是 BLOCK 对应的向量, 则标准化的向量 nf 为:

$$nf = \begin{cases} 1, & none-norm \\ \frac{v}{\|v\|_1 + \xi}, & L1-norm \\ \frac{v}{\sqrt{\|v\|_2^2 + \xi^2}}, & L2-norm \end{cases} \quad (9)$$

其中, ξ 为接近 0 的整数。最后将 BLOCK 对应的向量按照空间顺序排列形成 HOG 特征描述子。

(4) 依照上述方法将 $Object_{norm}$ 图像做灰度变换及 Re-size 处理, 提取模板图像的个性化感兴趣区域的 HOG 特征 HOG_{obj} 。

2.2.3 提取待检索图片特征

待检索图片中个性化感兴趣区域 $retrieve_{norm}$ 的定位与 HOG 特征提取。

(1) 感兴趣区域的粗定位。

归一化待检索图像, 以位置 $rect_{norm}$ 为中心, 扩展其区域 ($width * 3, height * 3$) 作为个性化特征的检测区域 Det_{rect1} 。

(2) 将模板图像的个性化区域 $Object_{norm}$ 与步骤(1)中的 Det_{rect1} 构建高斯金字塔。在 Det_{rect1} 区域中用式(10)归一化相关系数匹配法定位出与 $Object_{norm}$ 相似的个性化感兴趣区域 $retrieve_m$ 。

$$d_c(y_1, y_2) = \frac{y_1^T y_2}{\|y_1\| \|y_2\|} \quad (10)$$

(3) 提取 $retrieve_m$ 区域的 HOG 特征 HOG_{ret_m} , 其中 m 表示待检索图像的个数。

2.2.4 计算相似度

用式(1)计算 HOG 特征相似性。依据待检测图像与模板图像的个性化 ROI 的欧氏距离得出它们的相似度 $dist_m$ 。

2.3 基于全局及个性化感兴趣区域特征的检索算法

CNN 特征属于高层描述特征, 对个性化的局部特征描述能力较低。HOG 特征主要用来描述局部梯度分布特征, 用于

个性化感兴趣区域的特征匹配能够获得较好的局部检索效果。根据上述两种检索特征的优缺点, 本文提出了一种基于全局及个性化感兴趣区域特征的图像检索算法, 其综合了多个局部 HOG 特征与 CNN 全局特征, 提高了特征匹配的准确率。

$$dist_m = \alpha \lfloor \beta dist_{hog1} + (1-\beta) dist_{hog2} \rfloor + (1-\alpha) dist_{CNN} \quad (11)$$

其中, $dist_{hog1}, dist_{hog2}$ 表示待检测图像与模板图像的个性化 ROI1、ROI2 对应的 HOG 特征计算出的欧氏距离。 $0 \leq \beta \leq 1$ 是调节 ROI1 与 ROI2 特征欧式距离的组合因子。 $0 \leq \alpha \leq 1$ 是调节全局特征欧式距离与个性化 ROI 特征欧式距离的组合因子。如式(11)所示, 基于全局及个性化感兴趣区域特征计算得出的待检测图像与模板图像的相似度为 $dist_m$ 。

3 检索结果的评价

3.1 个性化检索的评价指标

本文利用图像检索领域中一些常用的评判标准^[13,14]对个性化检索结果进行评价。检索图像库含有 4 万张准卡口车辆图片, 其中包括 50 辆目标车辆, 每个目标车辆有 20 张相关图像。从检索结果中选取前 α 条作为 TOP α 集合, 将集合中与查询条件车辆为同一车辆的图像作为标准答案, 包括同一车辆不同时间的图像。针对 TOP α 集合进行评价, 常见的评价指标如下:

3.1.1 准确率(Precision Rate)

TOP α 中符合相关图像的记录条数与前 α 个结果的比例。准确率越高, 则检索效果越好。

3.1.2 检索率(Retrieval Rate)

TOP α 中符合相关图像的记录条数与所有相关图像的总条数的比例。

$$RR(q) = \frac{NF(\alpha, q)}{NG(q)} \quad (12)$$

其中, $NG(q)$ 表示数据集中与 q 主观相似的图像个数, $NF(\alpha, q)$ 表示在检索到的前 α 个结果中与 q 相似的图像个数。

3.1.3 平均准确率(Average Precision)

如式(13)所示:

$$AP = \frac{\sum_{i=1}^m \text{排在第 } i \text{ 个样本之间的相关图像条数} \times rel(i)}{\text{所有相关图像的总条数}} \quad (13)$$

$$rel(i) = \begin{cases} 0, & \text{第 } i \text{ 个样本不是相关图像} \\ 1, & \text{第 } i \text{ 个样本是相关图像} \end{cases}$$

其中, m 表示前 m 条检索结果。AP 值越高, 则相关图像越靠前。

4 检索结果

针对 TOP50 集合, 每个目标车辆对应 20 个相关车辆图像。在 CPU Intel i7-3770 的条件下, 根据检索的评价指标基于 CNN 特征的检索结果如表 2 所列, Compressed Net 模型的 fc6 层特征归一化后的检索结果检索率(RR)达到 87.1%, 平均准确率(AP)达到 78.66%。利用 HOG 特征的个性化感兴趣区域检索的结果如表 3 所列, 其对具有明显个性化感兴趣区域的图像检索效果比 CNN 全局特征好, 但对于个性化 ROI 不明显的图像的检索效果比较差。本文提出的基于全局及个性化感兴趣区域特征的图像检索算法的检索结果如表 3 所列, 检索率(RR)达到 94.32%, 平均准确率(AP)达到 90.45%。

(下转第 246 页)

文献[7]中算法有更好的检索效果。

结束语 本文在分析传统 GrabCut 算法的基础上,对其能量函数进行了优化,对其分水岭算法算子进行了改进,提出了改进的 GrabCut 算法,将改进后的 GrabCut 算法引入基于内容的服装图像检索系统中,实验结果表明:本文方法与同类方法相比,在检索显示准确性以及检索的平均查准率和查全率方面均有明显提升。

参 考 文 献

[1] Forczmański P, Czapiewski P, Frejlichowski D. Comparing Clothing Styles by Means of Computer Vision Methods[J]. Computer Vision and Graphics, 2014, 86(71):203-211

[2] Wu Xiao, Zhao Bo, Liang Ling-ling, et al. Clothing Extraction by Coarse Region Localization and Fine Foreground/Background Estimation[J]. Advances in Multimedia Modeling, 2013, 77(33):316-326

[3] 林瑶, 田捷. 医学图像分割方法综述[J]. 模式识别与人工智能, 2002, 15(2):192-204

[4] Mortensen E N, Barrett W A. Intelligent scissors for image composition[C]// Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques, 1995:191-198

[5] 郝凯. 图像及图像序列上的交互抠图技术研究[D]. 南京:南京理工大学, 2012

[6] Falcao A X, Udupa J K, Samarasekera S, et al. User-steered image segmentation paradigms: Live wire and live lane[J]. Graphical Models and Image Processing, 1998, 60(4):233-260

[7] Rother C, Kolmogrov V, Blake A. GrabCut: interactive foreground extraction using iterated graph cuts[J]. ACM Transactions on Graphics, 2004, 23(3):309-314

[8] Lee T S. Image representation using 2D Gabor wavelets[J]. Pattern Analysis and Machine Intelligence, 1996, 18(10):959-971

[9] Arifina A Z. Image segmentation by histogram thresholding using hierarchical cluster analysis[J]. Pattern Recognition Letters, 2006, 27(13):1515-1521

[10] 韦娜, 耿国华, 周明全. 基于内容的图像检索系统性能评价[J]. 中国图形图像学报, 2004, 9(11):1271-1276

(上接第 207 页)

表 2 CNN 特征信息及检索结果

模型特征	Compressed Net	Compressed Net	AlexNet
	Fc6	Fc6	Fc6
Norm	No	Yes	Yes
dim	2048	2048	4096
PR(%)	32.31	34.84	33.68
RR(%)	86.95	87.1	85.13
AP(%)	73.50	78.66	71.41
Time(s)	8.7	7.8	60

表 3 检索结果

模型特征	Compressed Net Fc6	HOG	Proposed
PR(%)	34.84	24.2	36.4
RR(%)	87.1	69.4	94.32
AP(%)	78.66	56.45	90.45
Time(s)	7.8	3	11

结束语 鉴于治安卡口、及准卡口图像数据的不断增加,待检索图片的数量也在不断飙升,如何在更大的数据集上进行快速有效的个性化图像检索,以及其中检索的精度、检索耗时的提升、跨平台的特征匹配等问题都有待研究并解决。因此,本文提出了一种基于全局及个性化感兴趣区域特征的图像检索算法,通过对 AlexNet 优化加速得到 Compressed Net 的 fc6 层特征并结合多个个性化感兴趣区域的 HOG 特征,利用欧氏距离计算出待检索图库中与模板图像相似度最高的相关图像。通过对 50 辆目标车辆图像的检索验证,结果表明,该算法的检索率、平均准确率都达到 90%,呈现出较好的检索效果,计算速度快,具有较强的鲁棒性和实用性。

参 考 文 献

[1] Smeulders A W M, Worring M, Santini S, et al. Content Based Image Retrieval at the End of the Early Years[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000, 22(12):1349-1380

[2] Zhou X S, Huang T S. Image Retrieval: Feature Primitives, Feature Representation, and Relevance Feedback[C]// Proceedings of IEEE Workshop on Content-based Access of Image and Video Libraries, 2000:10-14

[3] Breiteneder C, Eidenberger H. Content-Based Image Retrieval in Digital Libraries[C]// Proceedings of IEEE International Conference on Digital Libraries: Research and Practice, 2000:288-295

[4] Sun J D, Zhang X M, Cui J T, et al. Image Retrieval Based on Color Distribution Entropy[J]. Pattern Recognition Letters, 2006, 27(10):1122-1126

[5] Hinton G E, Qsindero S, Teh Y W. A Fast Learning Algorithm For Deep Belief Nets[J]. Neural Computation, 2006, 18(7):1527-1554

[6] Simonyan K, Zisserman A, Simonyan K, et al. Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Eprint Arxiv, 2014

[7] He K, Zhang X, Ren S, et al. Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(9):1904-1916

[8] Dong C, Chen C L, He K, et al. Learning a Deep Convolutional Network for Image Super-Resolution[C]// European Conference on Computer Vision, 2014:184-199

[9] Krizhevsky A, Sutskever I, Hinton G E. ImageNet Classification with Deep Convolutional Neural Networks[J]. Advances in Neural Information Processing Systems, 2012, 25(2):2012

[10] Lowe D. Distinctive Image features from scale-invariant key points[J]. International Journal of Computer Vision, 2004, 60(2):91-110

[11] Bay H, Tuytelaars T. SUFR: Speeded Up Robust Features[C]// Proc. of European Conference on Computer Vision, 2006:404-417

[12] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 2005:886-893

[13] Manjunath B S, Ohm J R, Vasudevan V V, et al. Color and texture descriptors[J]. IEEE Trans. Circ. Syst. Video Technol., 2001, 11(6):703-715

[14] Datta R, Joshi D, Li J, et al. Image Retrieval: Ideas, Influences, and Trends of the New Age[J]. ACM Computing Surveys, 2008, 40(2):1-60