

基于K中心算法分层视觉词典树构造方法

张星宇, 许钢, 张阳, 江娟娟, 符茂胜

(安徽工程大学检测技术与节能装置安徽省重点实验室, 安徽 芜湖 241000)

摘要:视觉词典树是通过视觉词袋(Bag of Visual Words, BoVW)模型将一组图像形成单词表征场景。它主要通过聚类算法构造视觉词典树,为了改进K-Means在寻找聚类中心的过程中受到噪声异常维度的干扰,基于K中心算法提出了一种分层视觉词典树构造方法。该方法提出了从每次迭代后的聚类特征点中随机地选取新的中心点,可避免因噪声异常维度使计算中心点位置时有较大误差,从而提高聚类质量,使类簇更加紧凑。在图像的分类实验中,分别使用K-Means和K中心算法对构建的视觉词典树进行性能评价,实验结果证明通过K中心算法构造的分层视觉词典树可有效提高图像分类准确率。

关键词:K-Means、K中心算法、视觉词典树、图像分类

中图分类号:TP181

文献标志码:A

引言

随着互联网技术的迅猛发展和图像资源日益丰富,信息检索、图像识别和分类变得越来越困难,如何从大型图像数据库中获得重要的信息是一个急需解决的问题。视觉词袋(Bag of Visual Words, BoVW)模型最初用于对文档进行处理,将文档拆分表示为一组无关的词,通过统计文档中的词频进行文档的相似匹配。如今,计算机视觉领域的研究人员用视觉词袋模型表征图像内容,用于机器人导航、图像语义建模和场景分类等方面^[1-3],它首先通过对一系列图像进行底层特征提取和描述,得到大量特征组成的特征空间,再对特征空间内的特征点进行无监督聚类形成K个类簇,然后采用分层聚类的方法对得到的K个类簇再次进行逐层聚类,直到

满足所要求的层数为止^[4],从而得到可以表征场景的视觉词典树,最后根据图像在视觉词典树上不同分布进行矢量表达。

词袋模型一般采用K-Means算法进行聚类,然而K-Means将类簇中特征点的平均值作为类簇中心,在聚类过程中,少量噪声等异常维度的数据就能对平均值产生极大的影响,导致计算得到的聚类中心与实际聚类中心位置有较大误差,从而使类簇失真^[5-6]。Kim s等人^[7]首先设置阈值,在进行聚类时去掉阈值以下的特征值,从而大大减少人为因素对物体形状和纹理的影响。Baek s等人^[8]通过建立任意维度观测数据的相似性矩阵,然后采用特征值分解的方法在低维空间实现聚类。Wang m等人^[9]通过均值漂移(Mean-Shift)分割方法,提取图像像素中的纹理特征,通过聚类形成视觉单词,

收稿日期:2017-09-18

基金项目:检测技术与节能装置安徽省重点实验室开放基金资助项目(2017070503B026)

作者简介:张星宇(1992-),女,安徽庐江人,硕士生,主要从事智能信息处理及应用方面的研究,(E-mail)1055109496@qq.com

希望产生高度有序并且语义强的特征,从而使系统整体的稳定性和高效性提高。根据上文的研究,本文提出了一种基于K中心算法视觉词典的构建方法,在聚类过程中,从每次迭代后的类簇特征点中随机地选取新的中心点,可避免因噪声异常维度使计算中心点时有较大误差,从而可提高聚类质量,使类簇更紧凑。

1 BoVW 表示图像的基本流程

视觉词袋(BoVW)是视觉词典的关键部分,视觉词袋是词袋(BoW)从文本检索领域向图像检索领域的一次自然推广^[10],可有效解决如今计算机视觉领域中海量的图像数据带来的问题。BoVW是用来表示场景图像^[11],得到图像关于视觉单词的矢量表示形式,如图1所示,其基本流程为:

(1) 根据不同应用选择适合的特征提取方法,使用该提取算法从图像数据集中提取特征点并进行描述。

(2) 将所有的特征点向量集合在一起形成特征空间,利用K中心算法将相似的特征点向量聚集在一起,形成K个类簇,通过计算每个类簇的中心位置并且用一个视觉单词来表示,由此得到一个包含K个单词的词典。

(3) 计算视觉词典中K个视觉单词的权重,根据图像在视觉词典树上的不同分布,从而将每个图像表示为K维数值向量。

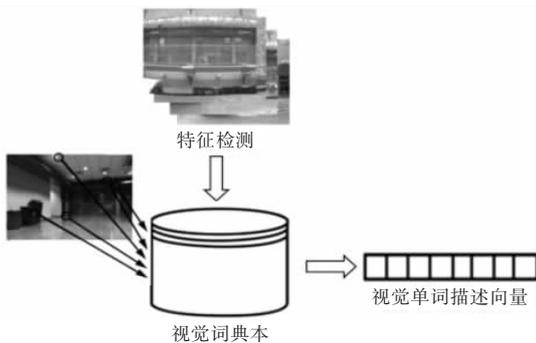


图1 BoVW 表示图像流程

2 基于K中心算法的视觉词典构造方法

在寻找中心点的过程中,K-Means会通过聚类形成K个类簇,对每个类簇中所有特征数据求取均值,并将此均值作为该类簇中心的位置^[12-13]。当聚类的特征

点中存在噪声时,少量噪声就会干扰计算中心点的过程,造成计算所得平均值和实际中心点之间有较大误差,从而使聚类结果失真。与K-Means不同的是,K中心算法从迭代产生的类簇中随机地选取一个特征点作为新的中心点。

2.1 K中心算法介绍

设 $X = \{x_1, \dots, x_n\} \in R^d$ 是从图像数据集采取到的 n 个特征数据,并将其组成的图像特征训练集; $C = \{c_1, \dots, c_k\}$ 是聚类中心向量集。

输入:簇的数目 k ,特征空间中 n 个特征点。

输出: k 个聚类中心。

(1) 随机选取 k 个中心点 $C = \{c_1, \dots, c_k\}$ 。

(2) 计算特征空间的各个特征点与各个中心点的距离,根据最近距离分割原理,将 n 个特征点划分到步骤(1)中选取的 k 个中心点为代表的类簇中。这里采用欧氏距离作为度量值:

$$dis(x, c) = \sqrt{\sum_{j=1}^d [x(j) - c(j)]^2} \quad (1)$$

(3) 随机选择一个非中心的特征点 c_{random} 。

(4) 计算用 c_{random} 代替 c_i 的绝对误差 E :

$$E = \sum_{i=1}^k \sum_{p \in X} |p - c_i|^2 \quad (2)$$

(5) 如果将步骤(3)中选取的特征点 c_{random} 作为新的中心点,其绝对误差能小于原中心点 c_i 所造成的绝对误差,那么该特征点则可取代原中心点。重复计算多个随机选取的特征点,选择绝对误差最小的特征点成为新的中心点,从而形成新 k 个中心点集合。

(6) 直到中心点不再发生变化。

根据上述算法可知, k 中心算法接受 n 个 d 维的特征向量作为输入量,然后将 n 个特征数据采用新的中心点选取方式划分为 k 个聚类,得到的 k 个聚类需满足:同一个聚类中的对象之间具有较高的相似度,而不同聚类中的对象差别较大。

2.2 分层视觉词典树构造

传统视觉词典一般采用平面结构形式,但这样不仅极大地限制了单词的数量,而且还降低了视觉词典的表征能力,增加了计算成本^[14]。视觉词典中包含的单词少,计算效率变高,但由于得到的视觉单词是通过将特征数据高度聚类的,其表征场景能力会下降,原本不同

的特征可能在视觉词典树上投射到同一个视觉单词;视觉词典中单词增多,则其表征场景能力提高,但计算负荷也会急剧增加。

为了解决上述问题,本文使用分层聚类算法构造分层视觉字典树^[15],分层结构可通过扩大视觉单词空间提高表征能力的同时降低计算复杂度。在训练阶段,从训练集的图像中提取特征并描述,将所有特征数据进行逐层聚类以生成视觉词典树。首先需要给定视觉词典树的分支数 K 和层数 L ,将提取到的所有图像特征数据通过 K 中心算法初步划分为 K 个分支,然后逐层地对每个分支再次执行聚类,得到下一层的 K 个分支,直到第 L 层停止,如图 2 所示,将每个分支用视觉单词来表示,从而形成视觉词典树。

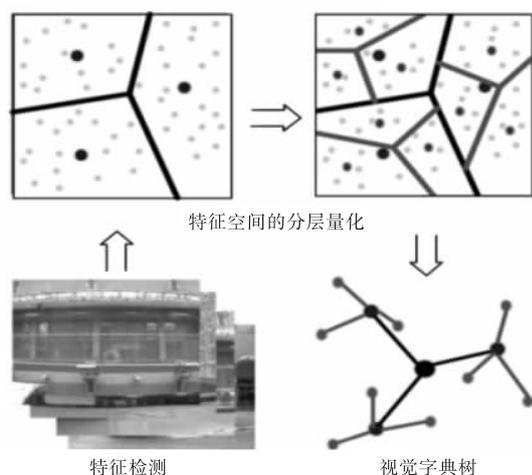


图 2 视觉词典树构造过程

词典树的中间节点仅供快速查询,只在第 L 层形成视觉单词,那么一棵 L 层 K 分支树生成的视觉单词个数为 K^L ,一棵 K 为 5, L 为 6 的视觉词典树就可以含有上万的视觉单词。在相同的计算复杂度的情况下,分层结构拥有的单词空间要远大于平面结构,因此分层结构可以较好地表征场景。

2.3 视觉单词重要性评估

在图像中提取出来的局部特征的重要性不同,有些频繁地出现在某些类别中而在其他类别中很少出现,同等对待视觉词典树节点的所有视觉单词,忽略不同的视觉单词在视觉词典树上的分布不同,造成视觉词典树不能正确表征场景^[16]。因此有必要对视觉单词进行加权,使得对类别区分越重要的视觉单词的权重越大,从

而提高 BoVW 模型在图像识别和分类中的效果。BoVW 模型最常使用的特征加权算法是 TF-IDF 算法^[17]。它的主要思想对给定图像进行特征提取,而词频 (term frequency, TF) 是某一特征在这个图像中出现的频率。对于在某一特定图像里的特征来说,它的重要性可表示为:

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \quad (3)$$

式中: $n_{i,j}$ 是图像 d_j 中的特征 t_i 数目,而分母则是图像 d_j 中包含的所有特征数量之和。

逆向文件频率 (inverse document frequency, IDF) 用来度量某一特征在图像数据集中重要程度的指标。该特征的 IDF, 是对图像总数量与包含该特征的图像数量的商取对数得到:

$$idf_i = \log \frac{|D|}{|\{j:t_i \in d_j\}|} \quad (4)$$

式中: $|D|$ 为图像集中所有图像的总数量, $|\{j:t_i \in d_j\}|$ 是包含特征 t_i 的图像数。

综合词频和逆向文件频率从而使真正有意义的词汇的权值更大,权重可表示为:

$$tf-idf_{i,j} = tf_{i,j} \times idf_i \quad (5)$$

3 实验过程及结果

在 MATLAB 环境,以图像分类为例对以上聚类方法构造的视觉词典树的性能进行评估。从通用图像数据集中选取 4 种不同类别的图像集共 360 幅,分为训练集 240 幅和测试集 120 幅。实验参数设置如下:

局部特征提取:利用 SIFT 算法对图像进行特征提取和描述。

视觉词典树构造:将训练集图像的 SIFT 特征组成训练特征集,设置不同的词典大小 K (本文实验取为 300、400、500 和 600),运用 K -Means 和 K 中心算法分别构造视觉词典树。

(1) 在实验中通过混淆矩阵 (Confusion Matrix)^[18] 的一些指标可以衡量算法的精度。矩阵的列表示预测类的实例,行表示实际类的实例,矩阵的主对角线为被分到正确类别的样本数,矩阵中剩余的元素为被分到错误类别的样本数。分别运用 K -Means 和 K 中心

算法构造视觉词典树的图像分类实验结果的混淆矩阵表示分别如图 3 与图 4 所示。运用 K - Means 和 K 中心算法构造视觉词典树在图像分类实验结果对比见表 1。

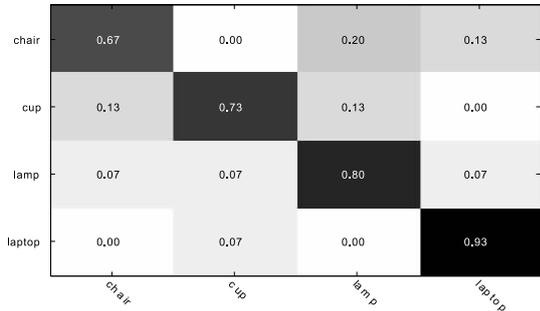


图 3 K - Means 聚类算法的分类结果混淆矩阵

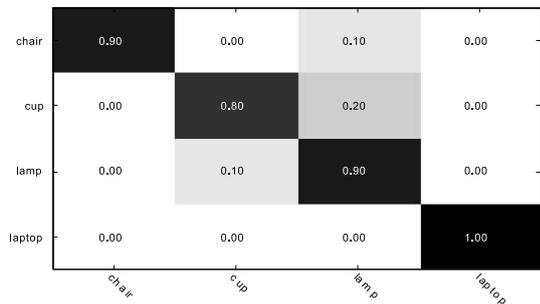


图 4 K 中心聚类算法的分类结果混淆矩阵

表 1 不同聚类算法在视觉词典树中分类结果对比

数据集	Chair %	Cup %	Lamp %	Laptop %	平均识别率 %
K - Means 算法	67%	73%	80%	93%	78.5%
K 中心算法	90%	80	90%	100%	90%

(2) 视觉词典中包含的单词少,计算效率变高,但由于得到的视觉单词是通过将特征数据高度聚类的,其表征场景能力会下降;视觉词典中单词增多,则其表征场景能力提高,但计算负荷也会急剧增加。设置聚类中心数即视觉单词数 K 为 300、400、500、600 和层数 $L=3$, 分别使用 K - Means 和 K 中心算法进行图像分类实验,如图 5 所示,其中横坐标表示设置的词典树中视觉单词的数量,纵坐标表示不同数量的视觉单词对应的分类准确率。

实验结果表明使用 K 中心算法形成的视觉词典树分类图像的准确率达到 90% 以上,K - Means 的准确率最高为 81.5%。通过增加视觉单词数量,即提高场景表征能力时,K - Means 和 K 中心算法的分类准确率都有

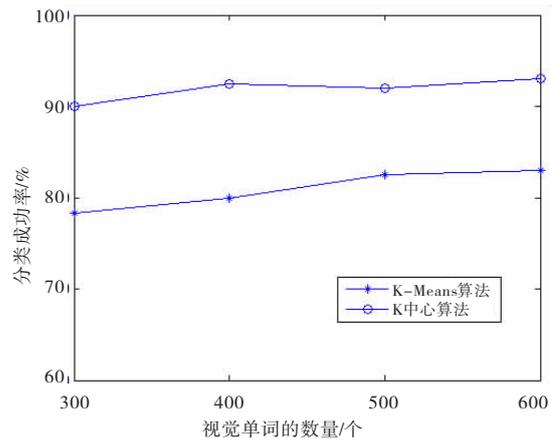


图 5 不同视觉单词数量在 K - Means 和 K 中心算法下的分类结果对比

所提高。

(3) 设置不同的层数 L 为 3、4、5 和 6, K 为 300, 分别使用 K - Means 和 K 中心算法进行图像分类实验,如图 6 所示,其中横坐标表示设置的词典树的层数,纵坐标表示不同层数的视觉词典树对应的分类准确率。

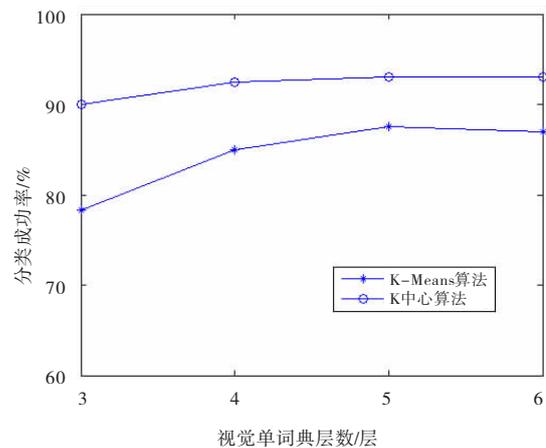


图 6 不同视觉词典树层数在 K - Means 和 K 中心算法下的分类结果对比

实验结果表明使用 K 中心算法形成的视觉词典树分类图像的准确率达到 90% 以上,K - Means 的准确率最高为 87%。随着视觉词典树层数的增加,图像分类的准确率趋于平稳,这是由于视觉词典树到达某一层数时就可以完整地表征场景。

4 结束语

本文对 BoVW 模型中基于 K 中心算法的无监督分层视觉词典构造方法进行研究分析,针对 K - Means 在

构造视觉词典树时对噪声敏感问题进行改进,提出了一种基于K中心算法的视觉词典构造方法。相比之下,K中心算法每次从迭代后的聚类特征点中随机地选取新的中心点,而选择的标准就是当该类簇中的特征点成为新的中心点后能提高类簇的聚类质量,使得类簇更紧凑,类簇的紧凑程度由误差函数确定。在通用图像库上开展的图像分类实验评估了K-Means和K中心算法视觉词典树的性能,实验结果证明K中心算法构造视觉词典树能有效提高图像分类准确率。不过,K中心算法在特征数据量大和需要构造较大的视觉词典时计算复杂度仍然较高,这是本研究工作下一步改进的方向。

参考文献:

- [1] WU L, HOI S C, YU N H. Semantics-preserving bag of words models and applications [J]. IEEE Transactions on Image Processing, 2010, 19(7): 1908-1920.
- [2] UIJLINGS J R, SMEULDERS A W, SCHA R J. Real-time visual concept classification [J]. IEEE Transactions on Multimedia, 2010, 12(7): 665-681.
- [3] 张永, 杨浩. 基于优化视觉词袋模型的图像分类方法 [J]. 计算机应用, 2017, 37(8): 2244-2247.
- [4] 孙勤红. 支持增量图数据的超图查询算法研究 [J]. 四川理工学院学报: 自然科学版, 2015, 28(3): 27-32.
- [5] 吴凤慧, 成颖, 郑彦宁, 等. K-means 算法研究综述 [J]. 现代图书情报技术, 2011(5): 28-35.
- [6] 盛靖友, 张洪伟. 基于预测强度的变量自动加权 K-Means 算法的研究与应用 [J]. 四川理工学院学报: 自然科学版, 2016, 29(2): 25-29.
- [7] KIM S, KWEON I S, LEE C W. Visual categorization robust to large intra-class variations using entropy-guided codebook [C] // Proceedings of 2007 IEEE International Conference on Robotics and Automation, Roma, Italy, April 10-14, 2007: 3793-3798.
- [8] BAEK S, YOO C D, YUN S. Learning a discriminative visual codebook using homonym scheme [C] // Proceedings of 2011 IEEE International Conference on Acoustics, Speech and Signal Processing, Prague, the Czech Republic, May 22-27, 2011: 2252-2255.
- [9] WANG M Y, ZHANG C L, SONG Y. Extraction of image semantic features with spatial mean shift clustering algorithm [C] // Proceedings of IEEE 10th International Conference on Signal Processing, Beijing, October 24-28, 2010: 906-909.
- [10] SIVIC J, ZISSERMAN A. Video Google: A text retrieval approach to object matching in videos [C] // Proceedings of the 9th IEEE International Conference on Computer Vision, Nice, France, October 13-16, 2003: 1470-1477.
- [11] LIU L, MA Y, ZHANG X, et al. High discriminative SIFT feature and feature pair selection to improve the bag of visual words model [J]. Int Image Processing, 2017, 11(11): 994-1001.
- [12] 孙可, 刘杰, 王学颖. K均值聚类算法初始质心选择的改进 [J]. 沈阳师范大学学报: 自然科学版, 2009, 27(4): 448-450.
- [13] GUPTA S, KUMAR R, LU K, et al. Local search methods for k-means with outliers [J]. Proceedings of the VLDB Endowment, 2017, 10(7): 757-768.
- [14] ZHANG J, ZHANG X Y, LEI Z, et al. An image classification method based on multiple visual dictionaries [C] // Proceedings of 2015 International Conference on Computer Science and Applications, Wuhan, November 20-22, 2017: 14-19.
- [15] PHILBIN J, CHUM O, ISARD M, et al. Object retrieval with large vocabularies and fast spatial matching [C] // Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, June 17-22, 2007: 1-8.
- [16] 赵刚. 基于视觉词袋模型的图像标注研究 [D]. 洛阳: 河南科技大学, 2012.
- [17] 施聪莺, 徐朝军, 杨晓江. TFIDF 算法研究综述 [J]. 计算机应用, 2009, 29(S1): 167-170, 180.
- [18] 米爱中, 张盼. 一种基于混淆矩阵的分类器选择方法 [J]. 河南理工大学学报: 自然科学版, 2017, 36(2): 116-121.

Visual Dictionary Tree Construction Method Based on K-Medoids

ZHANG Xingyu, XU Gang, ZHANG Yang, JIANG Juanjuan, FU Maosheng

(Anhui Key Laboratory of Detection Technology and Automation Devices, Anhui Ploytechnic University,
Wuhu 241000, China)

Abstract: Bag of Visual Words(BoVW) model is used in the visual dictionary tree to form a set of images to form visual words to represent the scene. It is mainly through the clustering algorithm to construct the visual dictionary tree, and in order to improve the K-Means interference noise in the process of finding abnormal dimension clustering center, a K-Medoids tree algorithm based on hierarchical visual dictionary is proposed. A new selection of the center from random clustering feature points after each iteration is provided, so the great error of the calculating center point caused by the abnormal noise dimension can be avoided, which can improve the quality of clustering, and the cluster is more compact. In the experiment of image classification, K-Medoids and K-Means are used respectively to evaluate the performance of the visual dictionary tree. The experimental results show that the hierarchical visual dictionary tree constructed by K-Medoids algorithm can effectively improve the accuracy of image classification.

Key words: K-Means; K-Medoids; visual dictionary tree; image classification

(上接第 28 页)

- [15] 任建新,边琦,张鹏,等.一种基于 SVD 和改进自适应算法的科氏流量计气体信号频率解算方法[J].传感技术学报,2013,26(10):1405-1408.
- [16] 杨辉跃,涂亚庆,张海涛,等.科氏流量计振动信号参数估计的最小二乘法[J].计算机工程与应用,2016,52(3):252-255.
- [17] 胡运权,郭耀煌.运筹学[M].3版.北京:清华大学出版社,2007.

Improvement of the Driving System of Coriolis Mass Flowmeter Based on Integral Filter and Nonlinear Fitting

XU Tongxu¹, HUANG Danping^{1,2,3}, YU Shaodong¹, GUO Kang¹, HOU Shanshan¹

- (1. School of Mechanical Engineering, Sichuan University of Science & Engineering, Zigong 643000, China;
2. Artificial Intelligence Key Laboratory of Sichuan Province, Zigong 643000, China;
3. Sichuan Provincial Key Lab of Process Equipment and Control, Zigong 643000, China)

Abstract: The drive circuit is the key factor whether the Coriolis mass flow meter can work normally. The flow pipe sometimes vibrates unsteadily due to random noise in the analog driving system and its completed oscillation cause much time. In order to decrease the start-oscillation time and increase vibration stability, an improved driving system, which based on integral filter and nonlinear fitting algorithm, has been put forward. In this improved driving system, the integral filter is used by calculating a summation of acquisition data, which has a better effect in decreasing noise. Then nonlinear fitting has been used to estimate the frequency of acquisition signal. The nonlinear fitting operating on the whole signal, as for its estimation will not be affected by the extra noise. The simulation results show that the improved algorithm has high accuracy than traditional algorithm. And experiments indicate that the improved driving system has a short time when the flow pipe vibrating stability, and improves the oscillation stability.

Key words: coriolis mass flow meter; improved driving circuit; integral filter; nonlinear fitting; composite signal