

文章编号: 2095-2163(2019)03-0097-04

中图分类号: TP391.41

文献标志码: A

一种改进词袋模型的图像分类算法

李咏豪

(南京理工大学 计算机科学与工程学院, 南京 210094)

摘要: 传统词袋模型易受视角、尺度和背景等因素干扰。本文对传统词袋模型进行了改进,提出一种基于角点检测和图论的感兴趣区域提取方法,以及结合高斯模糊隶属度函数选取视觉单词。首先,对图像进行角点检测,利用图论的方法划定 ROI 区域,然后对得到的 ROI 区域进行 SIFT 特征的提取并生成视觉词典,从而减少背景信息的影响。其次,引入高斯模糊隶属度函数改进图像视觉直方图的表示。在 Caltech 100 数据库上的实验结果表明,本文提出的方法相较于传统词袋模型,分类准确度提升了 3%。

关键词: 词袋模型; 角点提取; 图论; 高斯模糊隶属度函数

A novel image classification algorithm based on improved BoW model

LI Yonghao

(School of Computer Science and Engineering, Nanjing University of Science & Technology, Nanjing 210094, China)

[Abstract] The performance of the traditional Bag of Words (BoW) model is vulnerable to viewpoint, scale and background. This paper improves the traditional BoW model, and proposes the region of interest (ROI) extraction method based on the corner detection and graph theory. Moreover, the paper selects the visual words via Gaussian fuzzy membership function. First, use corner detection for images, and determine the ROI with graph theory; then, extract the SIFT features in the ROI and generate the visual dictionary to decrease the influence of the background information; finally, introduce the Gaussian fuzzy member function to improve the representation of visual histogram of image. The experimental results on Caltech 100 database show that, the accuracy of the proposed method is 3% higher than the traditional BoW model.

[Key words] Bag of Words; corner extraction; graph theory; Gaussian fuzzy membership function

0 引言

在计算机视觉领域中,图像分类是基础问题之一,被广泛应用于视频监控和图像分析等方面。图像分类中常用的方法是词袋模型^[1-3]。词袋模型可以分为以下 3 个步骤:

(1) 特征提取。利用 SIFT 算法^[4]生成 128 维的特征描述子,这些描述子具有尺度、光照等的不变性。

(2) 构造视觉词典。利用 k-means 聚类算法对 SIFT 描述子进行聚类,得到视觉单词,进一步构建视觉词典。

(3) 利用支持向量机(SVM)完成分类。

词袋模型分类效果易受视角、尺度、背景等因素干扰。针对这些问题,目前提出不少对于词袋模型的改进算法。Wang 等人^[5]提出空间金字塔匹配模型;Philbin 等人^[6]提出一种软分配的视觉词汇统计直方图的构建方法。同时,人类在识别图像时,往

往只是对某一区域感兴趣,而不是整幅图像。研究时,就可以通过对 ROI(感兴趣区域)进行特征提取,从而减少非感兴趣区域特征点的干扰。因此,本文通过对原图像进行角点检测,结合图论方法,确定 ROI,然后对 ROI 进行 SIFT 特征点提取,这样可以使提取的 SIFT 特征描述子集中在物体上,排除背景上的干扰点。另外,在视觉词汇直方图生成过程中,本文引入模糊隶属度函数来提高分类准确率。对此可做研究阐述如下。

1 ROI 区域定位

传统词袋模型是对整幅图像提取特征,如此一来,位于背景上的特征点也被提取出来,这将对视觉词典直方图产生一定影响。因此,本文采用角点和图论相结合的方法来减少背景特征点的影响。这里拟将探讨分述如下。

1.1 角点提取

Shi 等人^[7]提出用于追踪的 Shi-Tomasi 角点特

基金项目: 南京理工大学本科生科研训练“百千万”计划项目(201810288012)。

作者简介: 李咏豪(2000-),男,本科生,主要研究方向:模式识别、机器学习。

收稿日期: 2019-03-05

征,在视频跟踪领域有较好的效果。本文将 Shi-Tomasi 角点特征引入到图像分类中,和图论方法相结合,用于定位 ROI 区域。角点属于图像中的局部特征点,角点处的一阶导数为局部最大,并且图像的灰度值在水平与垂直方向上有一定变化。设图像在像素点 (x,y) 处的灰度值为 $I(x,y)$,以该点为中心构建一大小为 $n * n$ 的窗口 M ,窗口平移 (x,y) 后的灰度变化如下:

$$E(\Delta x, \Delta y) = \sum_{(x,y) \in M} \omega(x,y) [I(x + \Delta x, y + \Delta y) - I(x,y)]^2, \quad (1)$$

对式(1)进行泰勒展开,省略掉二次项及以上的项,得到 $E(\Delta x, \Delta y)$ 的矩阵表示,如式(2):

$$E(\Delta x, \Delta y) = [\Delta x, \Delta y] D \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix},$$

$$D = \sum_{(x,y) \in M} \omega(x,y) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}. \quad (2)$$

其中, I_x 与 I_y 表示图像在 x 和 y 方向的偏导数, $\omega(x,y)$ 为高斯滤波器。

矩阵 D 可以用来判定像素点是否为角点。设 λ_1 和 λ_2 是矩阵 D 的特征值,如果 $\lambda_1 > \lambda_2$,并且 $\lambda_2 > k\lambda_{2\max}$ ($\lambda_{2\max}$ 表示任意像素点较小特征值的最大值,本文设 k 值为0.05),则可判定此像素点为角点。图1列出提取角点的实例。



图1 角点提取

Fig. 1 Corner extraction

1.2 图论方法

本文利用图论^[8]思想,对提取的角点构建无向图,根据各角点的连续,去除关联度低的角点。这里,图 G 表示为 $G(V,E)$,其中 V 表示图中的顶点, E 表示图中各个点间的联系。研究中将1.1节中提取的角点作为图的顶点,然后通过式(3)构造无向图 M ,即:

$$M(i, j) = \begin{cases} d_{i,j}, & \text{if } d_{i,j} < \varepsilon; \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

其中, $d_{i,j}$ 表示顶点 i 与 j 间的距离, ε 表示阈值。

本文结合角点检测和图论来确定 ROI,具体步骤如下。

步骤1 角点检测及初始化 ROI。对各图像提

取 Shi-Tomasi 角点特征。以一幅图像为例,记录图像的角点 $Des_i(x,y)$, $i = 1, \dots, n$ 。初始化 ROI 区域,将整幅图像作为 ROI 区域,定义 ROI 区域的边界范围为 $x_{\min}, x_{\max}, y_{\min}$ 和 y_{\max} 。

步骤2 构建无向图。按式(3)构建角点的无向图 $M(i,j)$, $i, j = 1, \dots, n$, n 表示特征点的数目。

步骤3 去除关联度小的角点。按式(4)得到关键角点集合 V :

$$V = Des_i(x,y) \text{ if } \text{count}(M(i,:)) > \delta \quad (i = 1, \dots, n). \quad (4)$$

步骤4 划定 ROI 区域。根据步骤3得到关键角点的集合 V ,更新 $x_{\min}, x_{\max}, y_{\min}$ 和 y_{\max} ,划定 ROI 区域。图2显示了通过本文方法确定的 ROI。



图2 ROI定位

Fig. 2 The location of ROI

从图 2 可见,利用本文方法确定的 ROI 区域,可以有效地将前景与背景区分开,然后,在 ROI 提取 SIFT 特征描述子和生成视觉词典,这样可以减小背景特征点对视觉词汇的影响。

2 模糊理论的应用

传统词袋模型在视觉词典生成之后,在对每幅图像生成直方图时,是计算该幅图像中每一个特征向量与视觉词典中视觉词汇的距离,即采用硬分配方法,将其映射到距其最近的视觉词汇上面。如图 3 所示。图 3 中, A、B、C、D 是生成的视觉单词,1、2、3、4 是图像局部特征。在硬分配方式下,直接将特征 1 和 2 量化到词汇 A,特征 3 和 4 量化到词汇 B。但从图 3 中可以发现,特征 1、2、3 和 4 距离视觉词汇 A 和 B 的距离是不同的,因此,在生成图像直方图时对于视觉词汇的影响也应该不同。

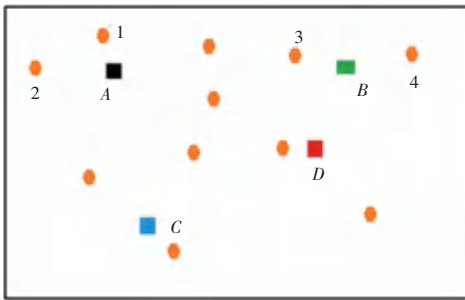


图 3 特征空间示意图

Fig. 3 Feature space diagram

这里,研究引入模糊隶属度函数^[9],考虑局部特征与视觉单词在特征空间的距离关系。设 $X = [x_1, x_2, \dots, x_N] \in R^{D \times N}$ 为一含有 N 个局部特征的集合, $C = [c_1, c_2, \dots, c_M] \in R^{D \times M}$, 为包含 M 个视觉词汇的视觉词汇表。对于局部特征 x_i , 在 C 上的响应为 $R_i = [r_1, r_2, \dots, r_M] \in R^{1 \times M}$, 其中,研究求得 R_i 的数学公式即如式(5)所示:

$$R_i = \arg \min_{R_i} \|x_i - CR_i^T\|_2^2, \text{ s.t. } \text{Card}(R_i) = 1, \|R_i\|_2 = f(R_i), \quad (5)$$

其中, $f(\cdot)$ 表示隶属度函数,本文利用高斯隶属度函数来提高分类准确率,如下式所述:

$$f(x, \sigma, \mu) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (6)$$

3 实验结果

本文拟在 Caltech-101 图像库上进行实验,该库中有从人到动物等 101 种类别,共包括有 9 146 幅图像,每个类别的图像数目从 31 到 800 张不等。

研究随机选择 10 种类别进行实验,包括:airplanes、anchor、ant、brain、ceiling_fan、chair、cup、elephant、face 和 rooster。每类随机提取 10 张作为训练图像,再随机选取 10 张和 20 张图像作为测试图像,进行 10 次重复实验,计算平均分类准确率。实验中,视觉词典的大小选为 300。

本文首先提取 ROI 区域,然后再计算分类准确率,分 2 组实验,一组用传统词袋模型,即对整幅图像提取 SIFT 特征;另一组是在 ROI 内进行 SIFT 特征提取,其中,参数 δ 为 2。实验结果见表 1。从表 1 可以发现,在 ROI 内提取特征,可以有效提高词袋模型的分精度。

表 1 分类准确度对比

Tab. 1 The comparison of classification accuracy %		
训练图像数目	传统方法	ROI 算法
10	51.3	53.5
20	53.9	56.3

接着,本文计算高斯隶属度函数对于图像分类精度的影响。在计算过程中,考虑高斯隶属度函数参数 $\sigma (\mu = 0)$ 对于分类精度的影响,如图 4 所示。从图 4 可见,当 $\sigma = [0.3, 0.5]$ 时,分类准确度达到最大值 54.9%。进一步,本文将 ROI 区域定位算法和高斯隶属度函数相结合,对其进行测试,所得结果见表 2。实验结果表明:相对于传统词袋模型,本文算法在分类准确度上大约提高了 3%。

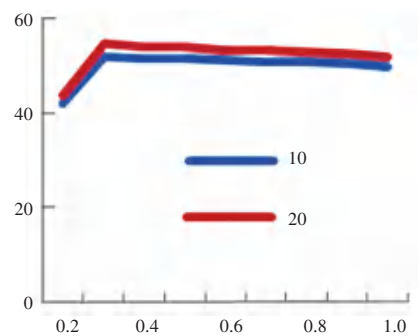


图 4 σ 对分类精度的影响

Fig. 4 The influence of classification accuracy with σ

表 2 分类准确度对比

Tab. 2 The comparison of classification accuracy %		
训练图像数目	传统方法	本文算法
10	51.3	54.4
20	53.9	57.5