

(1) 空圆特性: Delaunay 三角网是唯一的(任意 4 点不能共圆), 在 Delaunay 三角形网中任一三角形的外接圆范围内不会有其他点存在.

(2) 最大化最小角特性: 在散点集可能形成的三角剖分中, Delaunay 三角剖分所形成的三角形的最小角最大.

Delaunay 三角剖分具备如下优异特性.

(1) 最接近: 以最近邻的 3 点形成三角形, 且各线段(三角形的边)皆不相交.

(2) 唯一性: 不论从区域何处开始构建, 最终都将得到一致的结果.

(3) 最优性: 任意两个相邻三角形形成的凸四边形的对角线如果可以互换的话, 那么两个三角形 6 个内角中最小的角度不会变大.

(4) 最规则: 若将三角网中的每个三角形的最小角进行升序排列, 则 Delaunay 三角网的排列得到的数值最大.

(5) 区域性: 新增、删除、移动某一个顶点时只会影响临近的三角形.

(6) 具有凸多边形的外壳: 三角网最外层的边界形成一个凸多边形的外壳.

鉴于上述优点, Delaunay 三角剖分很适合描述点集的局部位置关系, 对于 n 个点, 计算时间复杂度为 $O(n \log n)$. 通过全局的位置限制, 得到匹配的视觉单词集, 基于这些视觉单词, 我们可以完成 Delaunay 三角剖分. 通过三角剖分来匹配邻近的点, 所有三角形内未匹配的点只能在其他图像的三角形内寻求匹配. 只要有新的匹配区域, 就可以继续进行 Delaunay 三角剖分, 在这一步, 其他的局部限制方法也可以加进来.

将计算几何的三角剖分方法引入视觉词袋模型, 通过对视觉单词集进行 Delaunay 三角剖分, 把图像空间上位置相近的视觉单词按照一定规则相连, 得到三角形网络. 如图 4 所示, 对视觉单词进行 Delaunay 三角剖分(包括图像的 4 个顶点), 将图像分成了多个三角形区域. 然后基于该网络寻找若干参考单词对, 进行匹配. DT 网络最大限度地避免了出现狭长、尖锐的三角形连接, 并具有唯一性. 当图像视觉单词表示给定时, 用 Delaunay 三角剖分对该点进行三角剖分时, 所产生的三角形网络是唯一的, 与视觉单词的排序无关, 只是与视觉单词的拓扑结构有关. 可以使得少数视觉单词的缺失、个别干扰视觉单词的出现和视觉单词定位偏差等因素对划分结果的影响尽量限定在一个小的局部范围内. 图 4(a) 中如果视觉单词 A 在图 4(b) 中没有对应点, 这表明单词 A 要么对于图 4(b) 来说是一种真正缺失的情况, 要么对于图 4(b) 来说是一个伪点, 但这都不影响在相应位置右下方找到两对相似三角形. 而对于图 4(a) 和图 4(b) 中对应的视觉单词 B 和视觉单词 C , 虽然视觉单词位置有偏差, 而这也影响到它们与周围若干视觉单词之间的结构, 但并不影响在视觉单词周围找到若干匹配的三角形对. 在视觉单词集中插入一个单词, 它仅仅影响到外接圆包含该插入点的三角形, 对 Delaunay 三角形网络只造成局部影响. 只要图像的一些区域没有严重损坏, 都可以进行准确定位, 从而有可能得以正确地识别.

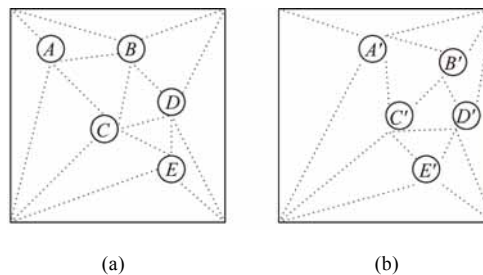


Fig.4 Delaunay triangle subdivision on visual word sets
图 4 视觉单词集进行 Delaunay 三角剖分

Delaunay 三角剖分还可以避免产生不稳定和瘦小的三角形. DT 网络具有极好的结构稳定性, 它保证了 Delaunay 三角剖分得到的三角形不是狭长的, 狭长的三角形会导致不稳定性甚至产生错误结果. 对几种我们熟知的拓扑结构进行比较后发现, Delaunay 三角剖分对于随机的位置扰动具有最好的结构稳定性.

2.5.2 基于三角形限制的图像匹配

两幅待匹配图像之间具有几何相似性约束,且 Delaunay 三角形网络可以唯一地表示其特征,所以可以利用两幅图像 Delaunay 三角网中三角形之间的相似度进行匹配.

三角形相似性匹配包括以下步骤.

1) 相似边

在两幅图像的 Delaunay 三角网络中,任意选择一个三角形的边,分别为 L_1 和 L_2 ,边长为 l_1, l_2 .边 L_1 连接视觉单词 w_{11} 和 w_{12} ,边 L_2 连接视觉单词 w_{21} 和 w_{22} .

判断两条边是否相似,需满足以下条件.

$$(1) |l_1 - l_2| < T_1.$$

$$(2) (w_{11} = w_{21}) \& (w_{12} = w_{22}) \vee (w_{11} = w_{22}) \& (w_{12} = w_{21}).$$

其中,阈值 T_1 根据实验情况设定,若以上条件满足,则 L_1 和 L_2 为相似边.如果不满足则从参考图像的 DT 网络中找下一条边和 L_1 进行匹配,直到图中所有三角边和 L_1 都比对过.如果所有边都比较过,且没有找到一对相似边,则表示两幅图像不匹配.

2) 相似三角形

在三角网络中,边的结构包含 4 个标志位 w_1, w_2, l, r . w_1 和 w_2 为两端的顶点单词, l 和 r 对应边的左右是否有三角形.根据一条边的左右标志,可以找到这条边的左右邻接三角形.现有如图 5 所示的 $Tri_1(l_{11}, l_{12}, \theta_{11}, \theta_{12}, w_1)$ 和 $Tri_2(l_{21}, l_{22}, \theta_{21}, \theta_{22}, w_2)$ 两个三角形, l_{11} 和 l_{12} 分别为视觉单词 P 与 w_{11} 和 w_{12} 的距离, l_{21} 和 l_{22} 分别为视觉单词 Q 与 w_{21} 和 w_{22} 的距离. $\theta_1, \theta_{11}, \theta_{12}$ 为 Tri_1 对应的 3 个夹角, $\theta_2, \theta_{21}, \theta_{22}$ 为 Tri_2 对应的 3 个夹角, w_1 和 w_2 为视觉单词类型.另外,基于三角形内角相似即可判断三角形相似的原理,可以引入三角形的模糊相似度判断^[29].三角形 Tri_1 的一个内角为 θ_1 ,那么在另一个三角形 Tri_2 的对应角与该角的差异性变化应该遵循下式的高斯分布:

$$d(\theta_2) = e^{-k(\theta_2 - \theta_1)^2} \tag{9}$$

这里, θ_2 表示另外一个三角形的对应角, $d(\theta_2)$ 表示这两个角的差异值.其中大于 0 的系数 k 可以由下式计算:

$$\begin{cases} k = \frac{1}{2 \times \sigma^2} \\ \sigma = \frac{\theta_1 \times P_{\theta_1}}{3} \end{cases} \tag{10}$$

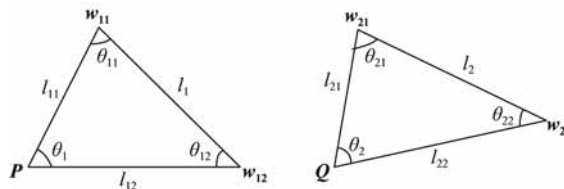


Fig.5 The comparison of similar triangle model
图 5 相似三角形的比较模型

式(10)是指 $\theta_1 \times P_{\theta_1}$ 就是高斯分布的 3σ 点,这里我们令 P_{θ_1} 为 50%.这样对角度的相似度可以由下式得到:

$$S_{\theta_1} = \cos^{n_s}(k_s(1 - d(\theta_2))) \tag{11}$$

基于余弦曲线的特征,我们可以定义 $n_s = 3, k_s = \frac{\pi}{2}$.对于一对三角形,可以先利用上式计算它们的内角相似度,然后利用下式计算整个三角形的相似度.

$$S_t = k_1 S_1 + k_2 S_2 + k_3 S_3 \tag{12}$$

这里, $k_i(i=1,2,3)$ 表示两个三角形相应内角的相似度权值,我们取 $k_1 = k_2 = k_3 = \frac{1}{3}$,也就是说,3 个内角对三角形相似度的影响是相同的.匹配相似三角形是通过模糊相似度阈值大于一个阈值 T_s 来获得的.

判断两个三角形是否相似,需要验证以下条件.

- (1) $|l_{11} - l_{21}| < T_1$ & $|l_{12} - l_{22}| < T_1$;
- (2) $|\theta_{11} - \theta_{21}| < T_2$ & $|\theta_{12} - \theta_{22}| < T_2$ & $|\theta_1 - \theta_2| < T_2$;
- (3) $w_1 = w_2$;
- (4) $S_i > T_3$,

其中, T_1, T_2 和 T_3 均是根据实验选取的阈值,相似三角形比较步骤如下.

Step 1. 选取一条相似边,如果有左右邻接三角形,则对其进行判断,若满足以上 4 个条件,则认为这两个三角形相似.

Step 2. 若不满足以上条件,则返回上一步骤,处理下一对相似边.

Step 3. 当所有的相似边处理完成,算法结束,统计得到的相似三角形并记录其对应的顶点单词.

3) 三角形内点的匹配

在全局拓扑限制下,还存在一些未匹配的视觉单词,这些视觉单词位于三角形内部.对于三角形内部点,只能在其他图像的对应三角形内部寻找匹配点^[27].

视觉单词限制在 Δabc 和 $\Delta a'b'c'$ 中,对于视觉单词 P_i 在 Δabc 中,它和 Δabc 三条边之间的关系为

$$P_i = \alpha + \beta(b-a) + \gamma(c-a) \quad (13)$$

β, γ 分别是 $(b-a)$ 和 $(c-a)$ 的尺度系数,已知 Δabc 的 3 条边 $a = [x_a, y_a, 1]^T, b = [x_b, y_b, 1]^T$ 和 $c = [x_c, y_c, 1]^T$. 参数 K 可以由下式计算:

$$K = \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} x_a & x_b & x_c \\ y_a & y_b & y_c \\ 1 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \quad (14)$$

其中, $P_i = [x_i, y_i, 1]^T, \alpha = 1 - \beta - \gamma$. 如果 Δabc 和 $\Delta a'b'c'$ 相似, P_i 和 $\Delta a'b'c'$ 中的单词 P_j 相匹配,则 $\Delta a'b'c'$ 的 3 条边也与参数 K 相关. 经过平移、尺度、旋转或者仿射变换后,估计的匹配单词 P_j 可以通过 P_i 的相关计算得到:

$$P_j = \begin{bmatrix} x_j \\ y_j \\ 1 \end{bmatrix} = \begin{bmatrix} x'_a & x'_b & x'_c \\ y'_a & y'_b & y'_c \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} \quad (15)$$

假定 Δabc 的 3 条边 $[x_a, y_a, 1]^T, [x_b, y_b, 1]^T$ 和 $[x_c, y_c, 1]^T$ 通过矩阵 H 变换为对应参照图像中 $\Delta a'b'c'$ 的 3 条边 $[x'_a, y'_a, 1]^T, [x'_b, y'_b, 1]^T$ 和 $[x'_c, y'_c, 1]^T$. 同样地, Δabc 内的视觉单词点 $[x_i, y_i, 1]^T$ 可以通过 $H[x_i, y_i, 1]^T$ 转换为 $[x'_i, y'_i, 1]^T$, 相关参数可以等价地表示为

$$K_H = \begin{bmatrix} \alpha' \\ \beta' \\ \gamma' \end{bmatrix} = \begin{bmatrix} x'_a & x'_b & x'_c \\ y'_a & y'_b & y'_c \\ 1 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} x'_i \\ y'_i \\ 1 \end{bmatrix} = \left(H \begin{bmatrix} x_a & x_b & x_c \\ y_a & y_b & y_c \\ 1 & 1 & 1 \end{bmatrix} \right)^{-1} \left(H \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} \right) \quad (16)$$

可以证明 $K_H = K$. 由于点的齐次坐标第 3 个维度保持为 1, 不需要进行归一化, 三角形约束具有平移、旋转、尺度和仿射变换不变性. 由于三角剖分将图像平面划分为很小的块, 所以对于噪声和扭曲更加鲁棒.

三角形内的视觉单词可能不止一个, 因此假设视觉单词集为 $w = \{w_j\}, P_j$ 位于单词集 w 中, P_i 和待匹配单词 w_j 的相似性得分由下式计算:

$$S_{i,j} = 1.5^{-(dist_{i,j}/R)^2} \times D_i^T D_{w_j}, \quad j = 1, 2, \dots, |w| \quad (17)$$

$dist_{i,j}$ 表示 P_i 和 w_j 的欧氏距离, D_i 和 D_{w_j} 分别表示 P_i 和 w_j 的视觉单词特征向量. 其中, $1.5^{-(dist_{i,j}/R)^2} \in (0, 1]$ 代表估计的待匹配单词 P_j 和单词 w_j 的空间距离, 点积 $D_i^T D_{w_j}$ 计算特征向量的相似性. 如果 w_j 的相似性得分高于一个预定的阈值 τ , 则将该单词视为 P_i 的一个匹配, 且 P_i 至多只有一个匹配, 阈值 τ 将在实验中选择.

通过对待匹配图像 P_A 的所有视觉单词进行处理后, 将会得到一个与 P_B 的匹配集合 T .

最终计算不同阶特征的核函数 $K(x,y)$,用于 SVM 训练.

3 实验结果与分析

实验在 Windows 7 操作系统(CPU 双核 2.60GHz,内存 4G),Visual Studio 2010 开发工具上由 C++语言实现.实验采用 Caltech101,COIL_100 和 Caltech256 这 3 个数据库验证本文算法.Caltech101 数据库包含 102 类共 9 145 幅图片,涵盖了由人到物的多种物体种类,每一类包含 40~800 张图片.COIL_100 数据库包含 100 种物体类别,每类物体包含 72 幅图片,类内图像具有复杂几何变换和仿射变换等特点.Caltech256 数据集包括 257 类共 30 607 幅图片,每个类别至少包含 80 张图片,相较于 Caltech101 数据集,物体种类更多,多样性增加.

我们采用的实验框架如下,首先进行图像分块与显著性区域提取,提取 128 维的 SIFT 特征描述子,使用标准的 K -means 聚类算法生成视觉单词表,其中,视觉单词容量参数 K 会在实验中讨论,然后结合单词的空间位置信息,构建视觉单词的空间几何模型,最后应用 SVM 进行图像分类.我们重复实验 10 次,每次随机选择训练图像集和测试图像集,获得平均准确率和偏差.每个类别随机选择多幅图像作为训练样本,剩下的用于测试.

在实验中,我们主要考虑影响分类性能的几个因素,包括显著区域的提取、分块数、视觉词典个数以及 SVM 核函数等.

首先,本文算法在提取特征之前进行显著性区域提取.应用显著区域提取方法,可以在训练集较小的情况下达到较好的分类性能,通过对图像进行显著区域提取,然后根据该区域进行特征提取和聚类,由此产生的视觉单词更具有代表性,能更精确地描述图像,而且能在一定程度上避免复杂背景信息和位置变化带来的干扰和影响.部分图像显著区域的提取效果如图 6 所示.传统方法在训练图像较少的情况下性能不太稳定,这是由于类内图像具有差异性所致.

接下来我们对分块数大小的选取进行分析,分块大小主要影响显著区域的大小和个数.选取分块大小为 4×4 、 5×5 、 10×10 ,分别计算不同分块大小对图像分类准确率的影响,分块过大,选取的显著性区域仍会包含大面积的背景区域.分块过小,使得包含一些显著度不高的目标区域被排除在外.在 Caltech101 数据集中将所有图像归一化为 200×200 ,分块数为 5×5 较好.



Fig.6 Extract the saliency region of image

图 6 图像显著区域提取

分析了显著性区域和分块数之后,我们将开始在 Caltech101 数据集上进行实验.在 Caltech101 数据集上选择 5 类物体,包括 airplane、euphonium、motorbike、camera、watch.训练集每类随机选择 30 张图像,测试集中每类随机选择 50 张图像.视觉词典大小为 500,实验重复 10 次,得到平均分类精确度,如图 7 所示.



Fig.7 Spatial geometry information modeling

图 7 空间几何信息建模

本文工作的主要目的是结合视觉单词的空间位置信息,以改进视觉词袋模型应用于图像分类中的性能.空间信息编码方法(SC)^[12]以及用户自定义感兴趣区域与空间分布相结合的方法(ROI+SL)^[13]均利用了图像特征的空间信息,因此,我们运用 SC 算法以及 ROI+SL 算法改进 BoW 用于图像分类实验中,并和本文算法进行比较,分析本文算法的有效性.

这部分我们针对视觉词典大小对分类性能的影响进行分析,这里采用不同大小的视觉词典 {50,100,200,300,500,1000}.如图 8 所示, x 轴表示词典大小, y 轴表示分类准确率,采用不同数目的视觉单词对分类性能产生一定的影响.根据分类准确率曲线,我们可以得出使用本文算法、SC 算法和 ROI+SL 算法的分类性能均有所提升,而且性能均好于传统的 BoW 模型.当视觉单词数目过小时,由于不同语义的图像块可能被标记为相似的视觉单词,使得无论是应用传统 BoW 模型还是改进的方法,它们的分类性能都相对较低.随着视觉单词数目的逐步增加,图像分类性能有所提升,由于传统 BoW 模型仅仅应用视觉单词的统计信息,忽略了视觉单词的空间信息,其分类准确率明显低于改进的方法,这说明,结合视觉单词的空间信息取得了效果.当视觉单词数目为 500 时,传统的 BoW 方法得到的分类准确率为 52.7%,SC 算法得到的分类准确率为 62.4%,ROI+SL 算法得到的分类准确率为 58.2%,而本文算法得到的分类准确率为 71.2%,加入空间信息后,分类准确率有很大的提升.

通过和 SC 算法以及 ROI+SL 算法的实验对比,可以看出本文算法在结合视觉单词空间位置信息上的优越性.SC 方法有效地编码了图像特征的空间位置关系,通过二进制的空间图描述特征对的相对空间位置,对局部特征施加几何限制.这种方法潜在要求待匹配图像具有重复的图像块,也就是说,具有相同或非常相似的匹配特征点的空间构型.由于存在无法避免的量化误差,该方法使用空间验证来消除错误的匹配.ROI+SL 算法结合用户自定义的感兴趣区域和空间分布进行特征匹配,基于用户定义感兴趣区域和图像分块的重叠百分比赋予图像特征不同的权重,图像的相似性计算通过单个图像块的相似性距离的线性组合计算得出.以上两种方法的分类准确率均低于本文算法,其中,SC 算法对于具有相同或非常相似的重复图像块的图像均有良好的分类效果,但是对于 Caltech101 数据库内类内差别较大的图像类别,分类效果提升得不会很明显,图像的仿射变换以及局部特征描述子的偏移也会对实验效果产生影响.ROI+SL 算法具有空间依赖性,用户选取的感兴趣区域对于算法效果影响显著,人工选取感兴趣区域虽然可靠性高,但是选取效率低,依赖于人机交互,过高的空间依赖导致图像分类效果提升有限.例如,一幅左下角有车的图像可能很难与右下角有车的图像分为同一类别.本文算法首先运用自适应的方法提取显著性区域,基于视觉单词拓扑结合策略,利用矩阵的特征分解有效地提取了关于视觉单词位置的全局信息.引入 Delaunay 三角剖分方法进行局部空间位置约束,图像空间上位置相近的视觉单词按照一定规则相连,得到 DT 网络,DT 网络具有良好的结构稳定性,可以使得少数视觉单词的缺失、个别干扰视觉单词的出现和视觉单词定位偏差等因素对分类结果的影响尽量限定在一个小的局部范围内.本文算法结合了一阶、二阶和三阶单词特征,将视觉单词的空间几何信息融入传统的视觉词袋模型,这里提到的 n 阶特征 ($n=1,2,3$),我们将在下文进行讨论.

本文算法不仅与 SC 算法和 ROI+SL 算法进行了比较,还与文献[30,31]中的图像分类实验结果进行了对比.文献[30]中主要通过对视觉单词的全局空间分布进行建模,并利用图像中相同视觉单词对的方向信息,改进传统 BoW 模型.文献[31]中提出空间金字塔匹配的改进方法,捕获更精确的图像特征空间信息.从实验结果可以看出,加入空间信息的图像特征表示方法具有较好的分类性能,其中,文献[30]的分类准确率为 67.1%,文献

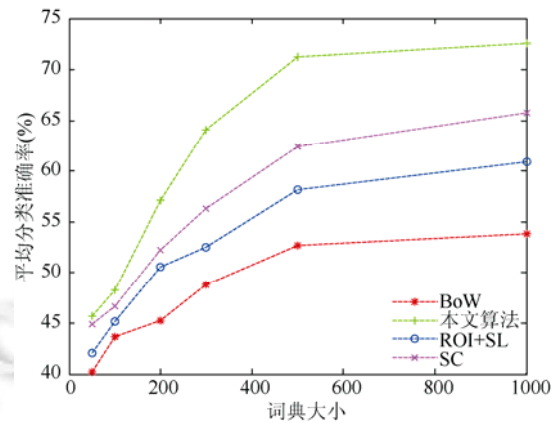


Fig.8 Dictionary capacity's impact on the average classification performance

图 8 词典容量对平均分类性能的影响

[31]的分类准确率为 65.93%,SC 算法的分类准确率为 62.4%,ROI+SL 算法的分类准确率为 58.2%.对比 BoW 的实验结果分别提升 14.4%,13.23%,9.7%和 5.5%.本文算法的分类准确率相比 BoW 提升了 18.5%,分类准确率高出以上其他方法,见表 1.

Table 1 Classification results for the Caltech101 dataset

表 1 Caltech101 图像库上的实验结果

方法	平均准确率/(%)
BOW	52.7
SC ^[12]	62.4
ROI+SL ^[13]	58.2
文献[30]	67.1
文献[31]	65.93
本文算法	71.2

在本文算法中,我们使用 Delaunay 三角剖分对单词的空间几何信息进行建模,几何信息建模的效果如图 7 所示.第 2.3 节已经介绍了高阶特征的定义,实验中,我们不仅单独使用 n 阶特征的核函数 $K_n(n=1,2,3)$ 进行分类实验,还使用 1 到 n 阶核函数的加权和去验证算法,其中, $u=0.02$.我们发现,无论是单独使用 n 阶特征的核函数还是进行加权和,从一阶(视觉单词)到二阶或三阶特征,效果都获得了提升.这表明,对视觉单词的空间几何信息建模起到了重要作用.然而随着阶数的增加,一些图像将不会出现共现的特征.分类效果的精确度将会随着我们提升阶数而有所提高(尽管会在某一阶数停止),我们发现虽然高阶核函数的精确度会有所下降,将它和其他阶特征结合在一起,但高阶特征仍能在一定程度上提升分类性能,事实上,高阶特征更具区分能力.

另外,随着阶数的上升,某些类别的图像类内相似性降低,原因在于某些类内图像具有更多变的目标结构、尺度和旋转,导致共现的高阶特征会减少,影响了分类性能.

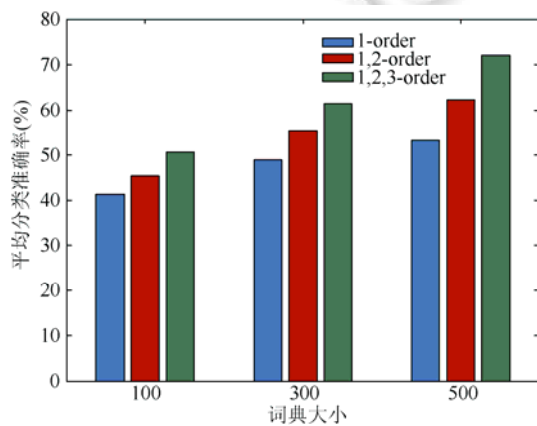


Fig.9 The influence of different order number for the average classification performance

图 9 不同阶数对于平均分类性能的影响

在分类过程中,我们应用 SVM 核函数,对于不同的核赋予不同的权重.如图 9 所示,增加特征阶数后,可以看到性能的提升.实验中,我们采用的词典大小为 500.我们尝试将词典大小降到 100 进行图像分类,在这种情况下,某些单个的视觉单词将可能变得无意义,在使用低阶特征进行分类时准确率很低.然而对于 SVM,当提高特征的阶数后,得到的性能非常接近我们使用 500 个视觉单词时所获得的性能.这表明,即使局部特征本身的区分能力不足,通过提高阶数也能够创造具有区别能力的特征.当将一阶、二阶和三阶特征结合时,可以达到最好的分类性能.

接下来将在 COIL_100 数据集上进行分类实验,在 COIL_100 数据集上随机选择 6 类物体,训练集中每类随机选择 20 张图像,测试集中每类随机选择 50 张图像.视觉词典大小为 500,实验重复 10 次,得到平均分类精确度.如图 10 所示,为 COIL_100 数据集上图像视觉单词空间几何信息建模结果.

本文实验还应用于 Caltech256 数据集上,我们在 Caltech256 数据集中选择 10 类物体,包括 butterfly、calculator、camel、necktie、elephant 等类别.训练集每类随机选择 30 张图像,测试集中每类随机选择 50 张图像.视觉词典大小为 1000,实验重复 10 次,得到平均分类精确度.

表 2 为 COIL_100 图像库中使用不同的方法进行图像分类的实验结果.将本文算法与传统的 BoW 模型以及 SC 算法^[12]、ROI+SL 算法^[13]和文献[32]中的方法进行了实验对比,可以看出,加入空间信息的特征表示显示了良好的分类性能,SC 算法编码了图像特征的空间位置关系,ROI+SL 算法基于用户自定义感兴趣区域,结合了图像特征进行了空间分布,均在一定程度上提升了图像特征的区分力和表达能力,提高了分类准确率.本文算法

较传统 BoW 模型的分类准确率提升 5.3%,效果最佳.表 3 为在 Caltech256 图像库中使用不同的特征表示方法进行图像分类的实验结果,Caltech256 图像库中类内变化更多样,分类难度有所增加.表中结果显示本文方法取得了较好的实验结果,与传统的 BoW 方法和空间金字塔匹配(SPM)方法^[2]、SC 算法、ROI+SL 算法相比,分类性能分别提高了 10.3%、7.4%、5.3%和 13%.文献[33]所采用的卷积神经网络方法效果最佳,平均准确率为 70.6%.然而,基于卷积神经网络的方法需要大量的样本和时间来学习特征,因此针对一些小的数据集,由于不需要预先的监督学习这一步骤,本文算法仍然能够起到重要的作用,效果更好.

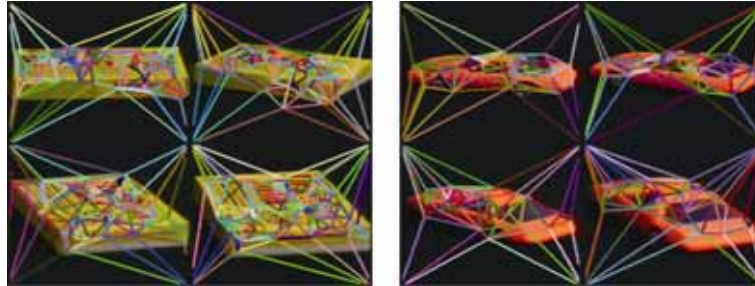


Fig.10 The spatial geometry information model of visual words

图 10 视觉单词空间几何信息建模

Table 2 Classification results for the COIL_100 dataset

表 2 COIL_100 图像库上的实验结果

方法	平均准确率(%)
BOW	91.3
SC ^[12]	93.2
ROI+SL ^[13]	91.6
文献[32]	90.52
本文算法	96.8

Table 3 Classification results for the Caltech256 dataset

表 3 Caltech256 图像库上的实验结果

方法	平均准确率(%)
BOW	31.2
SPM ^[2]	34.1
SC ^[12]	36.2
ROI+SL ^[13]	28.5
文献[33]	70.6
本文算法	41.5

本文算法主要针对传统视觉词袋模型进行改进,结合视觉单词的空间位置信息,提高图像特征表示的区分能力,并应用于图像分类.在采用词袋模型之前,使用显著性区域检测进行筛选,分类器是改进的 SVM.从表 1~表 3 中均可以看出本文算法取得了良好的效果,克服了图像的几何、旋转和仿射变换的影响,排除了复杂背景区域的干扰,对于类内差别较大的数据集,仍然能够避免背景干扰,有效应对类内变化,取得良好的分类效果.

4 总 结

视觉词袋模型广泛地应用于图像分类领域.在传统词袋模型中,视觉单词统计方法忽略了视觉词之间的空间信息以及分类对象的形状信息,导致图像特征表示区分能力不强.本文提出了一种改进的视觉词袋方法,结合显著区域提取和视觉单词空间位置结构,采用视觉单词拓扑结构策略和三角剖分方法,融入全局信息和局部信息,不仅能够产生更具代表性的视觉单词,而且能够在一定程度上避免复杂背景信息和位置变化带来的干扰.实验结果表明,本文算法将空间信息加入视觉词袋模型,并结合高阶特征,能够提高图像分类的精确度,而且与现有方法相比具有优越性.

我们还可以看到本文方法的改进之处,包括结合其他的核函数来改进视觉单词空间位置模型,一些其他的编码方法也可以引入,例如局部软分配方法^[34]可以用于软相似度加权.最后,空间信息还可以由多样的信息提供,例如颜色和纹理等,这是一个值得研究的方向.

References:

- [1] Sivic J, Zisserman A. Video Google: A text retrieval approach to object matching in videos. In: Proc. of the 9th IEEE Int'l Conf. on Computer Vision. IEEE, 2003. 1470-1477. [doi: 10.1109/ICCV.2003.1238663]

- [2] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Proc. of the 2006 IEEE Conf. on Computer Vision and Pattern Recognition. IEEE Computer Society, 2006,2:2169–2178. [doi: 10.1109/CVPR.2006.68]
- [3] Grauman K, Darrell T. The pyramid match kernel: Discriminative classification with sets of image features. In: Proc. of the 10th IEEE Int'l Conf. on Computer Vision, ICCV 2005. IEEE, 2005,2:1458–1465. [doi: 10.1109/ICCV.2005.239]
- [4] Lu ZW, Ip H. Combining context, consistency, and diversity cues for interactive image categorization. IEEE Trans. on Multimedia, 2010,12(3):194–203. [doi: 10.1109/TMM.2010.2041100]
- [5] Jiang Y, Wang RS, Wang C. Scene classification with context pyramid features. Journal of Computer-Aided Design and Computer Graphics, 2010,22(8):1366–1373 (in Chinese with English abstract).
- [6] Liu SY, Xu D, Feng SH, Liu D, Qiu ZD. A novel visual words definition algorithm of image patch based on contextual semantic information. Dianzi Xuebao, 2010,38(5):1156–1161 (in Chinese with English abstract).
- [7] Li T, Mei T, Kweon IS, Hua XS. Contextual bag-of-words for visual categorization. IEEE Trans. on Circuits and Systems for Video Technology, 2011,21(4):381–392. [doi: 10.1109/TCSVT.2010.2041828]
- [8] Zhang SL, Tian Q, Hua G, Huang QM, Li SP. Descriptive visual words and visual phrases for image applications. In: Proc. of the 17th ACM Int'l Conf. on Multimedia. ACM, 2009. 75–84. [doi: 10.1145/1631272.1631285]
- [9] Zheng QF, Wang WQ, Gao W. Effective and efficient object-based image retrieval using visual phrases. In: Proc. of the 14th Annual ACM Int'l Conf. on Multimedia. ACM, 2006. 77–80. [doi: 10.1145/1180639.1180664]
- [10] Sivic J, Zisserman A. Efficient visual search of videos cast as text retrieval. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2009,31(4):591–606. [doi: 10.1109/TPAMI.2008.111]
- [11] Wang MY, Zhang CL, Song Y. Extraction of image semantic features with spatial-range mean shift clustering algorithm. In: Proc. of the 10th IEEE Int'l Conf. on Signal Processing (ICSP). IEEE, 2010. 906–909. [doi: 10.1109/ICOSP.2010.5655732]
- [12] Zhou WG, Lu YJ, Li HQ, Song YB, Tian Q. Spatial coding for large scale partial-duplicate web image search. In: Proc. of the 18th ACM Int'l Conf. on Multimedia. ACM, 2010. 511–520. [doi: 10.1145/1873951.1874019]
- [13] Tian Q, Wu Y, Huang TS. Combine user defined region-of-interest and spatial layout for image retrieval. In: Proc. of the 2000 Int'l Conf. on Image. IEEE, 2000,3:746–749. [doi: 10.1109/ICIP.2000.899562]
- [14] Harel J, Koch C, Perona P. Graph-Based visual saliency. In: Advances in Neural Information Processing Systems. 2006. 545–552. <https://nips.cc/Conferences/2007>
- [15] Itti L, Koch C, Niebur E. A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. on Pattern Analysis & Machine Intelligence, 1998,(11):1254–1259. [doi: 10.1109/34.730558]
- [16] Hou XD, Zhang LQ. Saliency detection: A spectral residual approach. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, CVPR 2007. IEEE, 2007. 1–8. [doi: 10.1109/CVPR.2007.383267]
- [17] Ma YF, Zhang HJ. Contrast-Based image attention analysis by using fuzzy growing. In: Proc. of the 11th ACM Int'l Conf. on Multimedia. ACM, 2003. 374–381. [doi: 10.1145/957013.957094]
- [18] Achanta R, Hemami S, Estrada F, Susstrunk S. Frequency-Tuned salient region detection. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, CVPR 2009. IEEE, 2009. 1597–1604. [doi: 10.1109/CVPR.2009.5206596]
- [19] Li Y, Tsin YH, Genc Y, Kanade T. Object detection using 2D spatial ordering constraints. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, CVPR 2005. IEEE Computer Society, 2005,2:711–718. [doi: 10.1109/CVPR.2005.253]
- [20] Ma J, Ahuja N. Region correspondence by global configuration matching and progressive Delaunay triangulation. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. IEEE, 2000,2:637–642. [doi: 10.1109/CVPR.2000.854932]
- [21] Zeng D, Shi H, Zhang Q. Feature matching between images with multiple similar contents. Journal of Computer-Aided Design & Computer Graphics, 2011,23(10):1725–1733 (in Chinese with English abstract).
- [22] Jegou H, Douze M, Schmid C. Product quantization for nearest neighbor search. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2011,33(1):117–128. [doi: 10.1109/TPAMI.2010.57]
- [23] Kalantidis Y, Avrithis Y. Locally optimized product quantization for approximate nearest neighbor search. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2014. 2321–2328. [doi: 10.1109/CVPR.2014.298]
- [24] Wei SK, Xu D, Li XL, Zhao Y. Joint optimization toward effective and efficient image search. IEEE Trans. on Cybernetics, 2013, 43(6):2216–2227. [doi: 10.1109/TCYB.2013.2245890]

- [25] Zhang YM, Jia ZY, Chen T. Image retrieval with geometry-preserving visual phrases. In: Proc. of the 2011 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). IEEE, 2011. 809–816. [doi: 10.1109/CVPR.2011.5995528]
- [26] Zhang YM, Chen T. Efficient kernels for identifying unbounded-order spatial features. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, CVPR 2009. IEEE, 2009. 1762–1769. [doi: 10.1109/CVPR.2009.5206791]
- [27] Guo XJ, Cao XC. Good match exploration using triangle constraint. Pattern Recognition Letters, 2012,33(7):872–881. [doi: 10.1016/j.patrec.2011.08.021]
- [28] Tang G. Study on improved fingerprint identification algorithm based on Delaunay triangulation [Ph.D. Thesis]. Shenyang: Northeastern University, 2008 (in Chinese with English abstract). [doi: 10.7666/d.y1842528]
- [29] Li GH, Zhou DX, Dong L, Liu YH, Cai XP. Effective corner matching based on Delaunay triangulation. Signal Processing, 2007,23(5):695–698 (in Chinese with English abstract). [doi: 10.3969/j.issn.1003-0530.2007.05.012]
- [30] Khan R, Barat C, Muselet D, Ducottet C. Spatial orientations of visual word pairs to improve bag-of-visual-words model. In: Proc. of the British Machine Vision Conf. BMVA Press, 2012. 89.1–89.11. <http://bmvc2012.surrey.ac.uk/>
- [31] Zhang E, Mayo M. Improving bag-of-words model with spatial information. In: Proc. of the 25th Int'l Conf. of Image and Vision Computing, IVCNZ. IEEE, 2010. 1–8. [doi: 10.1109/IVCNZ.2010.6148795]
- [32] Han XH, Chen YW, Ruan X. Multilinear supervised neighborhood embedding of a local descriptor tensor for scene/object recognition. IEEE Trans. on Image Processing, 2012,21(3):1314–1326. [doi: 10.1109/TIP.2011.2168417]
- [33] Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: Proc. of the Computer Vision, ECCV 2014. Springer Int'l Publishing, 2014. 818–833. [doi: 10.1007/978-3-319-10590-1_53]
- [34] Liu LQ, Wang L, Liu XW. In defense of soft-assignment coding. In: Proc. of the 2011 IEEE Int'l Conf. on Computer Vision (ICCV). IEEE, 2011. 2486–2493. [doi: 10.1109/ICCV.2011.6126534]

附中文参考文献:

- [5] 江悦,王润生,王程.采用上下文金字塔特征的场景分类.计算机辅助设计与图形学学报,2010,22(8):1366–1373.
- [6] 刘硕研,须德,冯松鹤,刘镝,裘正定.一种基于上下文语义信息的图像块视觉单词生成算法.电子学报,2010,38(5):1156–1161.
- [21] 曾丹,史浩,张琦.多相似内容图像的特征匹配.计算机辅助设计与图形学学报,2011,23(10):1725–1733.
- [28] 唐果.基于 Delaunay 三角剖分的指纹识别改进算法的研究[博士学位论文].沈阳:东北大学,2008. [doi: 10.7666/d.y1842528]
- [29] 李赣华,周东祥,董黎,刘云辉,蔡宣平.基于 Delaunay 三角化的有效角点匹配算法.信号处理,2007,23(5):695–698. [doi: 10.3969/j.issn.1003-0530.2007.05.012]



汪荣贵(1966 -),男,安徽池州人,博士,教授,博士生导师,主要研究领域为视频大数据,多媒体技术,人工智能,模式识别.



丁凯(1992 -),男,硕士,主要研究领域为图像处理,视频大数据.



杨娟(1983 -),女,博士,讲师,主要研究领域为智能信息处理.



薛丽霞(1976 -),女,博士,副教授,主要研究领域为图像处理,视频大数据,多媒体技术.



张清杨(1990 -),男,硕士,主要研究领域为进化计算,图像处理,人工智能.