E-mail: jig@irsa.ac.cn
Website: www.cjig.cn
Tel: 010-64807995

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号: TP301.6　　**文献标识码:** A　　**文章编号:** 1006-8961(2013)09-1093-08
论文引用格式: 王科平,杨艺,王新良. 包空间多示例图像自动分类[J]. 中国图象图形学报,2013,18(9):1093-1100. [DOI: 10.11834/jig.20130905]

包空间多示例图像自动分类

王科平,杨艺,王新良
河南理工大学电气工程与自动化学院,焦作　454000

摘　要: 为了有效地解决多示例图像自动分类问题,提出一种将多示例图像转化为包空间的单示例描述方法。该方法将图像视为包,图像中的区域视为包中的示例,根据具有相同视觉区域的样本都会聚集成一簇,用聚类算法为每类图像确定其特有的"视觉词汇",并利用负包示例标注确定的这一信息指导典型"视觉词汇"的选择;然后根据得到的"视觉词汇"构造一个新的空间——包空间,利用基于视觉词汇定义的非线性函数将多个示例描述的图像映射到包空间的一个点,变为单示例描述;最后利用标准的支持向量机进行监督学习,实现图像自动分类。在 Corel 图像库的图像数据集上进行对比实验,实验结果表明该算法具有良好的图像分类性能。
关键词: 包空间;多示例学习;图像分类;视觉词汇

Automatic classification of multiple-instance image based on the bag space

Wang Keping, Yang Yi, Wang Xinliang
School of Electrical Engineering and Automation, Henan Polytechnic University, Jiaozuo 454000, *China*

Abstract: In order to effectively solve the multiple-instance image classification problem, we put forward a new classification method, which transforms the multiple-instance image into a single instance image in the new space-bag space. First, the whole image is regarded as a bag and each region as an instance of that bag. According to the same visual regions of image samples are put into one cluster and k-means clustering algorithm is used to determine the visual words for each class of images. At this step, we use the information that labels of negative samples are all known has been used to select the typical visual words. Then, we construct a new bag space with these visual words and use a nonlinear function based on these visual words to transform each multiple-instance image into a point in the bag space. Finally, standard SVMs are trained in the bag feature space to classify the images. Experimental results and comparisons on the Corel image set are given to illustrate the performance of the new method.
Key words: bag space; multiple-instance learning; image classification; visual words

0　引　言

图像自动分类问题是计算机视觉领域中备受关注的问题。近年来,很多研究者致力于这方面的研究,在该领域产生了一大批新的图像分类方法,图像分类效果也有了很大的改善[1-3]。图像分类方法是根据图像的内容将其划分到预定义类别的方法,是实现按语义内容来检索图像的一种重要方式[4]。图像语义往往由多个区域语义组合而成,如 beach

收稿日期: 2012-11-06;**修回日期:** 2013-03-02
基金项目: 国家自然科学基金青年基金项目(61100120);国家自然科学基金面上项目(41074090);河南理工大学博士基金项目(B2012-0670)
第一作者简介: 王科平(1976—),女,副教授,2011 年于北京邮电大学获信号与信息处理专业博士学位,主要研究方向为图像理解、模式识别。E-mail: wangkp@hpu.edu.cn

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS
Vol. 18, No. 9, Sep. 2013

类图像，一般都包含“sand”、“sky”、“water”、“tree”等语义区域，若将整幅图像视为包，图像中的每个区域看做是一个示例，则描述图像的是若干个示例的特征向量，因此图像分类问题常常被视为是多示例学习(MIL)问题。

1997 年 Dietterich 等人[5]提出了多示例学习框架，该框架主要用来预测药物分子的活性。多示例学习问世以来得到研究者极大的关注，并在短短几年内取得了一系列引人瞩目的理论成果和应用成果，被认为是与非监督学习、监督学习和强化学习并列的第 4 种机器学习框架。多示例学习主要用于药物分子预测、图像检索、股票选择和文本分类等领域。该文主要是研究多示例学习在图像分类中的应用。

近 10 年间，很多多示例学习算法被提出[6-13]，并被广泛地应用到图像检索、自动图像分类、标注中。多示例学习算法与有监督学习算法是不一样的，在标准的监督学习算法中，训练集中的每个样本或者示例都是有确切类别标记的，而在多示例学习算法的训练集中，每个包有确切的类别标注，但是包中示例的标注是不确定的，负包示例都为负，而正包示例有正有负。为了能够借鉴成熟的监督学习的算法解决多示例学习问题，研究者们提出了很多相关的算法。针对图像分类或图像检索应用方面，这些算法大致可以分为如下两类：一类是根据多示例学习问题的特性，将已有的机器学习算法进行扩展，使其能够应用于多示例学习问题的处理。如 Andrews 等人[6]将多示例学习与支持向量机相结合，提出的两种基于支持向量机 SVM 的 MIL 算法，MI-SVM 和 mi-SVM，这两种算法首先根据多示例问题的特点，将标准的 SVM 算法进行扩展，最后通过启发式迭代算法最大化分类间隔来解决分类问题。其中 MI-SVM 主要是从包的角度进行分析，可以实现包的分类；而 mi-SVM 主要是从示例的层面进行考虑的，可以实现示例的分类。由于这两种算法的目标函数是非凸的，算法中采用迭代求解的方法常常会导致无法得到全局最优解，Gehler 等人[7]提出了 AL-SVM 算法，采用确定性退火的方法来解决上述非凸优化问题。另外，路晶等人[8]在 mi-SVM 算法的基础上提出了启发式 SVM 多示例学习，该算法在 mi-SVM 算法的每次迭代中，通过改变其中一个样例的类别来最大化普通 SVM 的分类间隔，在图像分类中，取得了很不错的分类效果。基于非对称支持向量机的多示例学习算法(ASVM-MIL)[9]也是通过扩展 SVM 算法来实现多示例图像分类的一种方法。ASVM-MIL 通过对错分的正例和负例引入不同的惩罚函数来将 SVM 应用于多示例图像分类。

上述算法都需要将已有的机器学习算法进行扩展，转化为能够解决多示例问题的形式，而实际上，很多已有的机器学习算法本身就具有很好的学习性能，所以，出现了第 2 类多示例图像分类算法。该类算法是通过改变多示例学习问题自身，使其可以用已有的、成熟的机器学习算法进行处理。Chen 等人通过空间转化的方法先后提出了两种经典的算法：DD-SVM[10]和 MILES[11]。其共同思想是：从图像中提取典型示例构造一个目标空间，然后将所有图像通过非线性函数映射到目标空间的一个点，使每幅图像在新空间实现单示例描述，这样就可以直接用标准的支持向量机实现图像的自动分类。DD-SVM 算法由于采用多样性密度(DD)算法求解图像中的典型示例，所以非常费时，而且对噪声比较敏感。MILES 算法对该算法进行了改进，采用训练包中所有示例构建目标空间，但是这种方法导致目标空间维数非常高，存在大量的冗余信息。李大湘等人[12]提出了基于半监督多示例学习的图像检索算法，与 DD-SVM 不同的是该算法采用“点密度”最大的原则来选择图像示例构建包空间，该算法可以有效提高构建包空间的实时性。文献[13]提出了 EMD-CkNN 算法，该算法采用推土机距离来度量图像之间的相似程度，由于推土机距离允许图像区域之间多对多匹配，所以可以很好地反映图像的整体相似度，而且该算法还可以降低图像分割误差对分类的影响。

在基于包层的多示例图像分类中，不同的典型示例构建方法会取得不同的分类效果，本文尝试提出一种新的示例构造方法，用于将多示例描述的图像分类问题转化到包层面去解决。

具体的做法是：在训练集中，描述图像的语义关键词是已知的，也就是说样本的标注是已知的。如果能够确定表示图像的单一视觉特征向量，则可以直接应用标准的监督学习算法实现图像的分类，因此，问题就转化为如何将多示例描述的图像转化为可以用单一特征向量来描述。根据具有相同视觉区域的样本都会聚集成一簇，利用聚类算法为每类图像确定其特有的视觉词汇，然后由这些视觉词汇构建一个新空间——包空间。在选择视觉词汇时，直

接采用聚类算法确定的视觉词汇会存在大量的冗余,所以本文采用负包中示例标注全部为负这一确定信息来消除区分度不高的冗余视觉词汇。最后利用非线性函数将多个示例描述的图像映射到包空间的一个点,变为单示例描述,并直接采用标准的支持向量机进行监督学习,实现图像自动分类。

针对多示例图像分类算法,探索新的示例构造方法是个非常值得研究的关键问题,本文所提示例构造的主要思想包括以下两个方面:

1) 通过对图像区域分布情况的分析发现,每类图像都存在部分典型的、特有的视觉区域,如果能找到这些视觉区域,就可以确定该类图像的语义类别。DD-SVM [10]、RSTSVM-MIL [12] 等算法都是基于整个训练集来构建典型示例的,而本文是针对不同的类别分别为每类图像确定其特有的视觉词汇,这种方法得到的视觉词汇能够更加全面地描述每个图像类别。

2) 在视觉词汇选择过程中,发现通过聚类算法得到的视觉词汇并非都是该类图像的典型代表区域,其中部分视觉词汇属于冗余词汇。冗余词汇主要存在于以下两种情况: 一是部分视觉词汇出现的频率很低,它们不具有典型的代表性,则把这类区域视为噪声; 二是在正、负两类中都会频繁出现的视觉区域,这种类型的视觉区域的类别区分度较低,所以不把它们作为代表图像类别的典型视觉词汇。针对上述问题,提出采用负包中示例标注都为负这一确定信息来指导“典型视觉”词汇的选择。

1　本文算法

1.1　问题描述

图像分类问题并不适合被视为普通的监督学习问题,因为在训练数据集中,图像的类别标签或标注关键词通常是针对整幅图像的,而不是关联具体的某个图像区域。也就是说,训练集中没有图像区域与标注关键词具有确切对应关系的数据,如果直接采用监督学习将会导致较差的分类效果。而图像训练集的这种情况与多示例学习问题非常类似,图像分割的区域可以视为多示例学习问题中的样例,整幅图像视为一个包,若标注为正包说明该图像的区域中至少有一个与类别标签或关键词相对应的区域,否则为负包,这样可以将图像分类问题视为多示例学习问题。

为了将图像分类问题转化为适合传统 SVM 分类问题,需要将多示例描述的图像映射到包空间,在图像包的层次上进行分类预测。在基于包层多示例学习的图像分类算法中,首先需要构建一个新的空间——包空间; 然后将多示例图像从示例空间映射到包空间,在包空间每幅图像对应空间的一个点; 当图像集中每一幅图像都可以用包空间的一个特征向量来表示时,就可以采用传统的支持向量机算法来实现图像自动分类。

1.2　构建包空间

在训练集中,包的标注是已知的,每个包都包含若干个示例,首先需要找出每个包的典型示例,即图像的典型视觉区域——视觉词汇。在同一类图像中,具有相同视觉特征的区域都会聚集成一簇,如图 1 所示,所以利用聚类算法对训练集中同一类图像区域聚类,利用聚类中心来表示该类图像所包含的视觉区域。用视觉词汇来表示每个视觉区域。通过聚类,可以得到该类别中所有的视觉词汇。

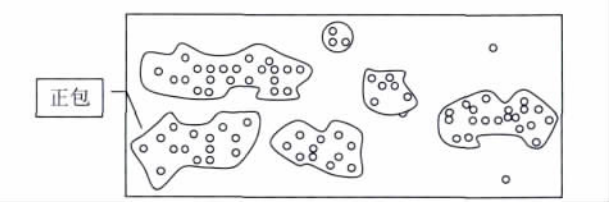

图 1　正包示例的分布

Fig. 1　Samples distribution of the positive bags

然而通过聚类处理后,得到的视觉词汇数量比较大,导致表示图像的特征向量的维数也大。由于向量的维数大,会使得分类效率、运行的速度降低,浪费资源。通过对图像区域的分析发现,聚类后会产生大量的对分类没有帮助的 0 噪声词汇,由于向量空间模型的工作模式是计算向量之间的相似度来确定待分类图像应该归到哪一类别中,因此这样会影响分类效果,降低分类精度。进一步分析发现,聚类处理后,每个视觉词汇对图像分类的贡献是不同的,多个类别都普遍存在的视觉词汇对分类的贡献小,在某特定类别中出现的频率大而在其他类别中出现频率小的视觉词汇对图像分类的贡献大。为了提高分类精度,应去除那些在多个类别中通用的,对于每一类别表现力不强的词汇,筛选出针对该类的典型视觉词汇特征向量集合。

以 Corel 图像库中 20 类图像中的 beach 和 elephant 两类图像为例,对其中的视觉词汇进行分析。

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS
Vol. 18, No. 9, Sep. 2013

如图 2，第 1 行图像属于 beach 类，该类图像包含的视觉词汇对应的语义内容有“beach”、“sky”、“sea”、“tree”、“people”、“mountain”等。图 2 第 2 行中的 elephant类，该类图像包含的视觉词汇对应的语义内容有“elephant”、“grass”、“sky”、“tree”、“water”等。

在 Corel 图像库中，这两个类别图像分别包含 100 幅图像，对其中每个视觉词汇出现的概率进行统计，统计结果见表 1。

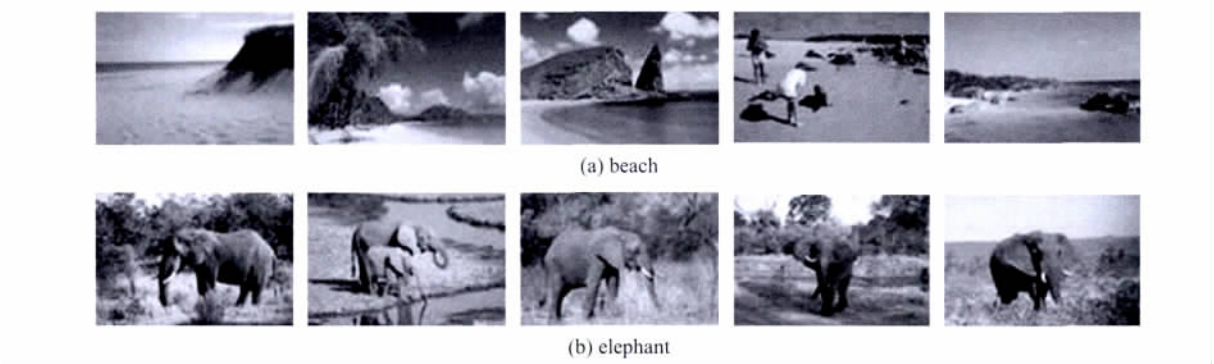

图 2　beach 类和 elephant 类中的图像

Fig. 2　Image simples of beach class and elephant class

表 1　beach 类和 elephant 类中所包含视觉词汇的数量

Table 1　The number of visual vocabulary in beach class and elephant class

类	beach	sea	people	mountain	tree	sky	elephant	grass	water
beach	91	89	80	51	10	94	0	0	0
elephant	0	0	0	0	78	90	100	94	15

表 1 纵向表示两类图像，横向表示每类中所包含视觉词汇的个数。由表 1 可以看出类中有些视觉词汇出现的概率是非常高的，如 beach 类中的“beach”、“sea”、“people”；elephant 类中的“elephant”、“tree”、“grass”等，这些词汇几乎在每一幅图像中出现，它们是构成各自语义类别的基本部分，是各自类别中的典型词汇。而有些视觉词汇出现的概率则较低，如 beach 类中的“tree”，只在 10 幅图像中有明确的体现；同样的情况出现在 elephant 中的“water”，这些视觉区域出现的概率要远远小于上述词汇，它们对类别的贡献相对要低很多。

另外，从表 1 可以看出，“sky”这个视觉词汇在两个类别中出现的概率都很高，它属于在多个类别中都普遍存在的视觉词汇，类别区分度较低，对分类的贡献小。

通过上述分析，可以看出冗余词汇主要存在于以下两种情况：一是部分视觉词汇出现的频率很低，如 beach 类中的“tree”，elephant 类别中的“water”，它们不具有典型的代表性，则把这个类别区域视为噪声；二是在正、负两类中都会频繁出现的视觉区域，如 sky，这种类型的视觉词汇类别区分度低，所以不把它们作为代表图像的典型视觉词汇。这一思想与文本分类中的特征选择类似，如果某个词在多个类别的文本中出现，则说明这个词的区分度不高，不选择作为文本分类的特征。

下面是典型视觉词汇的具体选择算法。

假设训练集 $\boldsymbol{B}=\{(\boldsymbol{B}_1,y_1),(\boldsymbol{B}_2,y_2),\cdots,(\boldsymbol{B}_l,y_l)\}$ 为已知类别的图像集，共 l 幅图像。其中 $\boldsymbol{B}_i$ 表示已标注的图像。图像集中每幅图像 $\boldsymbol{B}_i$ 被分割为若干个区域，$\boldsymbol{B}_{ij}$ 表示第 i 幅图像中的第 j 个示例（图像区域）。y_i 表示图像的标注，取值为 $+1$ 或 -1，若 $y_i=+1$ 表示该图像是属于某个语义概念类别，$y_i=-1$ 则表示图像不属于该语义类别，设 $\boldsymbol{S}^+=\{x_k^+ \mid k=1,2,\cdots,n\}$ 表示训练集中所有标注 $y_i=+1$ 的示例排在一起组成的集合，$\boldsymbol{S}^-=\{x_k^- \mid k=1,2,\cdots,m\}$ 则表示训练集中所有标注 $y_i=-1$ 的示例排在一起组成的集合，其中 n 和 m 分别表示训练集中正、负示例的个数。具体算法步骤如下：

1）对集合 $\boldsymbol{S}^+$ 中的示例进行 k-means 聚类。随机选取 k 个图像区域作为初始聚类中心。

$$k=\max N_j \tag{1}$$

式中，N_j 表示第 j 个正包中示例的个数。如果选择

10 个正包的话,则 $j=1,2,\cdots,10$。

2) 计算训练示例中每个区域与这些聚类中心的距离,并且根据最小距离进行相应对象的划分。

$$d(x_i^+,c_j)=\sqrt{\sum_{l=1}^{L}(x_{il}^+-c_{jl})^2} \tag{2}$$

式中,x_i^+ 为待聚类的示例,c_j 为第 j 个类别的聚类中心,L 表示示例的维数,x_{il}^+、c_{jl} 分别表示 x_i^+ 和 c_j 的第 l 个特征。在所有距离结果中,如果 $d(x_i^+,c_j)$ 最小,则把第 i 个示例划分到第 j 类中。

3) 重新计算每个聚类的中心

$$c_j'=\frac{1}{n_j}\sum_{i=1}^{n_j}x_i^j \tag{3}$$

式中,n_j 为划分到 j 类的示例个数,x_i^j 为第 j 类的示例。

4) 循环步骤 2) 3) 直到类别中心不发生变化或目标准则函数 E 最小化。

$$E=\sum_{i=1}^{N}\sum_{j=1}^{k}d(x_i,c_j) \tag{4}$$

式中,x_i 表示训练集中第 i 个示例。

5) 通过聚类可得到 k 个类别,计算每个类别的类别中心

$$c_j=\frac{1}{N_j}\sum_{i=1}^{N_j}x_{ij} \tag{5}$$

式中,c_j 表示第 j 类的类别中心,N_j 表示通过聚类后第 j 类所包含示例的个数,x_{ij} 表示第 j 类的第 i 个示例。

6) 聚类后,考察每个类别包含的示例个数,如果类中包含的示例个数很少,即

$$N_j'<n_0 \tag{6}$$

则把这个类别作为噪声去掉。

7) 计算保留下来的每个正包示例聚类中心 c_j 分别与每个负包示例 x_i^- 的距离,若

$$d=\|c_j-x_i^-\|>d_0 \tag{7}$$
$$j=1,2,\cdots,k;\ i=1,2,\cdots,m$$

则 c_j 选择作为该类的典型视觉词汇。式(7)表示正包示例聚类中心 c_j 与负包示例 x_i^- 的距离越小,则认为该类型图像与负包中某些示例相似性较高,它作为特征的区分度不高,所以不选择作为代表该类的典型视觉词汇。阈值 d_0 取为

$$d_0=\max\|c_j-x_i^j\|,j=1,2,\cdots,k;\ i\in N_j' \tag{8}$$

8) 选出同时满足式(9)两个条件的正包示例聚类中心作为包空间的基本特征。

$$\begin{cases}N_s'>n_0\\ \|c_j-x_i^-\|>d_0,j=1,2,\cdots,k;\ i=1,2,\cdots,m\end{cases} \tag{9}$$

式中,N_s'表示某个类别所包含示例的个数,n_0 为一个阈值。

1.3 包特征计算

设 $V=\{v_1,v_2,\cdots,v_c\}$ 是由 1.2 算法得到的 C 个区分度较强的典型视觉词汇,由这些典型的视觉词汇构建一个新的特征空间,称为包空间。将所有的图像映射到该空间,从而在包空间得到每幅图像的一个包特征向量。

图像 $B_i=\{B_{ij}\mid j=1,2,\cdots,N_i\}$ 在给定视觉词汇 V 下的特征为

$$\begin{cases}\phi(B_i)=[s(v_1,B_i),s(v_2,B_i),\cdots,s(v_c,B_i)]\\ s(v_k,B_i)=\min\limits_{j=1,2,\cdots N_i}\|B_{ij}-v_k\|\end{cases} \tag{10}$$

式中,$\phi(B_i)$ 为包 B_i 在包空间的特征,$B_i=\{B_{ij}\mid j=1,2,\cdots,N_i\}$ 中 B_{ij}表示包 B_i 中的示例,N_i 表示包 B_i 中共有 N_i 个示例。式(10)计算的是图像包 B_i 中每个示例分别与视觉词汇的最小距离,也就是计算包中示例与每个视觉词汇的相似程度。

所有图像都被分割为若干个区域,每类图像都有自己典型的代表区域,通过聚类算法选择出代表每类图像的典型视觉词汇,即得到具有很好区分度的 $V=\{v_1,v_2,\cdots,v_c\}$。通过式(10)的计算把多示例表达的图像转换到包空间的一个点,每幅图像就可以用一个特征向量来表示,这时多示例学习问题就转化成了典型的监督学习问题,可以用监督学习算法来实现图像的自动分类。

1.4 基于支持向量机的图像自动分类

支持向量机结构简单、具有全局最优性和较好的泛化能力,是求解模式识别和函数估计问题的有效工具,也是典型的有监督学习算法之一。监督学习要求训练数据集中样本与标注的模式是明确的一一对应关系,而多示例问题中每个有标注的样本都由多个示例描述,所以无法直接将支持向量机用于多示例学习问题。通过 1.2 节、1.3 节算法,将多示例图像嵌入成包空间的一个点,变成了单示例学习问题,则可以用标准的 SVM 进行监督学习。

设训练集由 l 个样本对构成的数据集 $T=\{(\phi(B_1);y_1),(\phi(B_2);y_2),\cdots,(\phi(B_l);y_l)\}$,其中 $\phi(B_i)$ 是由前边 1.2 节、1.3 节推导得到的图像

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS
Vol. 18, No. 9, Sep. 2013

在包空间的特征向量，$y_i \in \{+1, -1\}$，SVM 的目标是设计一个超平面，将所有的训练样本正确分类。具有最大分类间隔的最优超平面，可以使分类误差最小，所以最优分类面的构造转化为求解

$$\min_{w,b,\xi} \frac{1}{2} \| \boldsymbol{w} \|^2 + C\sum_{i=1}^{l} \xi_i$$
$$\text{s. t.} \quad y_i [w\phi(\boldsymbol{B}_i) + b] \geqslant 1 - \xi_i \qquad (11)$$
$$\xi_i \geqslant 0; i = 1, 2, \cdots, l$$

式中，参数 C 为错误分类的惩罚因子，ξ_i 是松弛变量。

采用拉格朗日乘子法求解这个具有线性约束的二次规划问题，则式(11)转化为

$$\max \frac{1}{2}\sum_{i,j=1}^{l} a_i a_j y_i y_j K(\phi(\boldsymbol{B}_i), \phi(\boldsymbol{B}_j)) - \sum_{i=1}^{l} a_i$$
$$\text{s. t.} \quad \sum_{i=1}^{l} a_i y_i = 0 \qquad (12)$$
$$0 \leqslant a_i \leqslant C; i = 1, \cdots, l$$

式中，$K(\phi(\boldsymbol{B}_i), \phi(\boldsymbol{B}_j)) = \varphi(\phi(\boldsymbol{B}_i)) \cdot \varphi(\phi(\boldsymbol{B}_j))$ 是核函数，a_i 是拉格朗日乘子。式(12)是求解不等式约束下的二次规划问题，根据最优化理论中的 KTT 条件，可得最优分类判别函数为

$$f(\boldsymbol{B}) = \text{sgn}\left(\sum_{i=1}^{l} a_i y_i K(\phi(\boldsymbol{B}_i), \phi(\boldsymbol{B}_j)) + b\right) \qquad (13)$$

当 $f(\boldsymbol{B})$ 取值为 +1 时，包 $\boldsymbol{B}$ 标注为正；$f(\boldsymbol{B})$ 取值为 −1 时，包 $\boldsymbol{B}$ 标注为负。

2 实 验

实验使用的图像数据集由 Corel 图像库中的 2 000幅图像组成，包括 20 个类别，每类 100 幅图像。为每类图像定义一个语义关键词来描述这类图像的内容，分别为：African、beach、buildings、buses、dinosaurs、elephants、flowers、horses、mountains、food、dogs、lizard、models、sunsets、cars、waterfall、antiques、battle ships、skiing and dessert。图 3 是从这 20 个类别中每一类随机选择的一幅图像。本文采用 Chen 等人在文献[10]中提出的图像自动分割算法和底层特征抽取技术，这些底层特征包括：3 维的小波纹理特征、3 维的 LUV 颜色特征以及 3 维形状特征，即每个图像区域的底层特征由一个 9 维的向量组成。

称前 10 类的 1 000 幅图像为“Corel 1k 库”，整个 2 000 幅图像为“Corel 2k 库”。在“Corel 1k 库”上，将包含 1 000 幅图像的 10 类图像随机平均分为两部分，一部分作为测试数据，一部分作为训练数据。从训练集的某一类图像中随机选出 10 幅图像作为正包，从其他 9 类图像中选择 10 幅图像作为负包，用文中 1.2 节所提算法求解第 1 类图像的典型区域，即视觉词汇。在选择每类图像的典型区域时，需要对 10 个正包中的示例区域进行 k-means 聚类。重复上述操作，选出 10 类图像各自的典型区域代表。实验中，通过设定 n_0 可以辅助选择典型的视觉

图 3 20 类图像的示例

Fig. 3 Image samples from the 20 classes

词汇,该参数的取值不能太大,否则可能导致丢失有用信息。选择的取值方法如下: n_0 取值为正包数的 1/4,即如果聚类后,某个类别的示例数量小于 1/4 正包个数,则表示该类区域在正包中出现的次数比较少,不是该类图像的典型区域,不选该类的中心作为视觉词汇。

为了验证算法的有效性,实验是在与 DD-SVM 算法同等条件下进行的。表 2 和表 3 分别给出了本文算法和 DD-SVM 在 10 类中具体的分类情况。表中对角线上的数据为该类的平均精确度,也就是正确分类的百分率,非对角线上的数据则表示其他类别误分到某一类的百分率。这里列出的是 5 次实验的平均结果。(表 2 和表 3 中的 0 ~ 9 个数字分别代表 African、beach、buildings、buses、dinosaurs、elephants、flowers、horses、mountains and food 这 10 个语义类别)。

在"Corel 2k 库"上,每类图像随机选取出一半用于训练,剩下的图像作为测试集,和其他的基于机器学习的多示例学习算法进行对比实验,10 次随机划分重复实验的平均分类准确率见表 4。

表 2　DD-SVM 算法在 10 类图像中的平均分类结果

Table 2　The average classification result of DD-SVM method on 10 class images

/%

类别	0	1	2	3	4	5	6	7	8	9
0	**67.7**	3.7	5.7	0.0	0.3	8.7	5.0	1.3	0.3	7.3
1	1.0	**68.4**	4.3	4.3	0.0	3.0	1.3	1.0	15.0	1.7
2	5.7	5.0	**74.3**	2.0	0.0	3.3	0.7	0.0	6.7	2.3
3	0.3	3.7	1.7	**90.3**	0.0	0.0	0.0	0.0	1.3	2.7
4	0.0	0.0	0.0	0.0	**99.7**	0.0	0.0	0.0	0.0	0.3
5	5.7	3.3	6.3	0.3	0.0	**76.0**	0.7	4.7	2.3	0.7
6	3.3	0.0	0.0	0.0	0.0	1.7	**88.3**	2.3	0.7	3.7
7	2.3	0.3	0.0	0.0	0.0	2.0	1.0	**93.4**	0.7	0.3
8	0.3	15.7	5.0	1.0	0.0	4.3	1.0	0.7	**70.3**	1.7
9	3.3	1.0	0.0	3.0	0.7	1.3	1.0	2.7	0.0	**87.0**

表 3　本文算法在 10 类图像中的平均分类结果

Table 3　The average classification result of our method on 10 class images

/%

类别	0	1	2	3	4	5	6	7	8	9
0	**78.0**	2.8	4.1	0.3	0.0	5.6	3.5	1.4	0.1	4.2
1	3.1	**69.2**	5.2	2.8	0	4.8	1.3	3.5	8.4	1.7
2	5.3	3.1	**76.9**	1.3	0	3.7	0.7	1	5.4	2.6
3	0.5	3.6	0	**92.5**	0	0	0	0	0	3.4
4	0.0	0.0	0.6	0.0	**99.0**	0.0	0.0	0.0	0.0	0.4
5	5.2	2.1	6.8	0.7	0	**76.4**	1.3	4.8	1.8	0.9
6	3.6	0	0	0	0	1.1	**89.8**	2.0	0.3	3.2
7	2.1	0.6	0	0.3	0	1.7	0.6	**93.2**	1.1	0.4
8	1.3	12.4	3.8	0.6	0	4.3	1.3	1.1	**72.7**	2.5
9	3.0	1.0	0	2.2	0.7	1.7	1.4	1.6	0	**88.4**

从实验数据可以看出,本文算法的平均分类精确度高于算法 DD-SVM,尤其是 African 这类图像,比 DD-SVM 高出 10.3%。从这类图像包含的视觉内容可以看出,很多图像区域中并没有很明确的语义视觉概念,分割出的区域比较多,这些区域在图像中的重要程度相差不大,如果只选出多样性密度最大值(DD)作为该类图像的特征代表,会丢掉该类很多其他的重要信息。而本文算法可以将该类中所有的有区分度的特征选出,从而大大提高了该类别的识别率。

中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS
Vol. 18, No. 9, Sep. 2013

表 4 列出了本文算法和其他 4 种算法分类精确度的对比值，这 4 种算法分别是 DD-SVM 算法、文献[13]的 EMD-CkNN 算法、文献[11]提出的 MILES 算法和由 Andrews 等人提出的 MI-SVM 算法，这 4 种对比算法是比较典型的用已有的机器学习算法来解决多示例问题的方法。

由表 4 的实验结果可以看出，不论是在“Corel 1k 库”数据库上还是在“Corel 2k 库”数据库上，本文算法的平均分类精确度都高于其他多示例图像分类算法。

表 4　5 种算法分类精确度的对比

Table 4　The classification accuracy contrast of five methods

	平均分类精度/%	
	Corel 1k 库	Corel 2k 库
本文算法	83.6	68.3
DD-SVM	81.5	67.5
EMD-CkNN	82.0	68.2
MILES	82.6	68.7
MI-SVM	74.7	54.6

3　结　论

图像自动分类问题是计算机视觉领域中备受关注的问题，由于图像底层特征与高层语义概念之间“语义鸿沟”的存在，目前已有的自动图像分类方法性能还不能令人满意。本文从多示例学习框架角度出发实现图像自动分类问题，提出了将多示例图像转化为单示例描述，然后利用监督学习 SVM 来进行图像分类的算法。由于该算法选择的视觉词汇可以较好地表达图像中所包含对象的语义，所以由视觉词汇构建的包空间具有较强的图像描述能力，利于提高图像标注的精度。在 Corel 图像数据集上的实验结果表明本文所提算法具有优良的图像分类性能，尤其适用当数据集中图像内容相对复杂、图像分割区域较多的分类与识别。在今后的工作中，将继续研究典型视觉词汇的提取规模，以及将本文所提算法应用于大规模图像集的自动分类。

参考文献(References)

[1] Wang J, Yang J, Yu K, et al. Locality-constrained linear coding for image classification [C] //Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, California: IEEE Press, 2010: 3360-3367.

[2] Zhou X, Yu K, Zhang T. Image classification using super-vector coding of local image descriptors [C] //Proceedings of the 11th European Conference on Computer Vision. Heraklion, Crete, Greece: Springer Press, 2010: 141-154.

[3] Tang Y J, Xu D, Xie W J, et al. A novel image scene classification method based on category topic simplex [J]. Journal of Image and Graphics, 2010, 15(7): 1067-1073. [唐颖军，须德，解文杰，等. 一种基于类主题空间的图像场景分类方法 [J]. 中国图象图形学报, 2010, 15(7): 1067-1073.]

[4] Liu Y, Zhang D, Lu G, et al. A survey of content-based image retrieval with high level semantics [J]. Pattern Recognition, 2007, 40: 262-282.

[5] Dietterich T G, Lathrop R H, Lozano-Pérez T. Solving the multiple instance problem with axis-parallel rectangles [J]. Artif. Intell, 1997, 89(1-2): 31-71.

[6] Andrews S, Hofmann T, Tsochantaridis I. Multiple instance learning with generalized support vector machines [C] //Proceedings of the 18th National Conference on Artificial Intelligence. Edmonton, Canada: AAAI Press, 2002, 943-944.

[7] Gehler PV, Chapelle O. Deterministic annealing for multiple-instance learning [J]. Journal of Machine Learning Research, 2007, 2: 123-130.

[8] Lu J, Ma S P. Region-based image annotation using heuristic support vector machine in multiple-instance learning [J]. Journal of Computer Research and Development, 2009, 46(5): 864-871. [路晶，马少平. 使用基于多例学习的启发式 SVM 算法的图像自动标注 [J]. 计算机研究与发展, 2009, 46(5): 864-871.]

[9] Yang C, Dong M, Hua J. Region-based image annotation using asymmetrical support vector machine-based multiple instance learning [C] //Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE Press, 2006, 2: 2057-2063.

[10] Chen Y X, Wang James Z. Image categorization by learning and reasoning with regions [J]. Journal of Machine Learning Research, 2004, 5(8): 913-939.

[11] Chen Y X, Bi J B, Wang J Z. MILES: multiple-instance learning via embedded instance selection [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2006, 28(12): 1931-1947.

[12] Li D X, Peng J Y, Li Z. Object-based image retrieval using semi-supervised mil algorithm [J]. Control and Decision, 2010, 25(7): 981-986. [李大湘，彭进业，李展. 基于半监督多示例学习的对象图像检索 [J]. 控制与决策, 2010, 25(7): 981-986.]

[13] Li D X, Peng J Y, He J F. Image categorization based on emd-cknn multi-instance learning algorithm [J]. Journal of Optoelectronics. Laser, 2010, 21(2): 303-306. [李大湘，彭进业，贺静芳. 基于 EMD-CkNN 多示例学习算法的图像分类 [J]. 光电子. 激光, 2010, 21(2): 303-306.]