

一种结合波段分组特征和形态学特征的高光谱图像分类方法

张帆¹ 杜博² 张良培¹ 张乐飞²

(武汉大学测绘遥感信息工程国家重点实验室 武汉 430079)¹ (武汉大学计算机学院 武汉 430072)²

摘要 如何准确识别图像中的类别信息,是计算机视觉和模式识别领域的重要研究问题。遥感卫星图像数据,尤其是高光谱等遥感图像数据的出现,将空间信息与光谱信息集成于同一数据集中,丰富了图像信息来源。如何准确地识别高光谱图像中的地物类别,已经成为了图像处理和模式识别领域的热点问题。面向高光谱图像数据提出了一种基于波段分组特征和形态学特征的高光谱图像分类方法,结合空间和光谱特征提高分类精度。通过真实的高光谱数据实验证明:利用波段分组可以有效地保持光谱特征,降低数据冗余;在波段分组基础上结合形态学特征进行分类,比传统分类方法的分类精度明显提高。

关键词 分类,高光谱图像,特征选择,形态学,支持向量机

中图分类号 TP751 文献标识码 A DOI 10.11896/j.issn.1002-137X.2014.12.059

Band Grouping Based Hyperspectral Image Classification Using Mathematical Morphology and Support Vector Machines

ZHANG Fan¹ DU Bo² ZHANG Liang-pe¹ ZHANG Le-fei¹

(State Key Laboratory for Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China)¹

(Department of Computer, Wuhan University, Wuhan 430072, China)²

Abstract How to analysis and recognize image accurately is an important issue in computer vision and pattern recognition fields. Remote sensing image, especially hyperspectral images combine spatial and spectral information in one data cube. In this paper, we proposed a band grouping feature selection method, then extracted morphology features. A feature selection algorithm called recursive feature elimination was applied to decrease the dimensionality of the input morphology features data. A support vector machine was used for the final classification. Experiments performed on real hyperspectral images, confirm that it is efficient using band grouping and mathematical morphology.

Keywords Classification, Hyperspectral image, Features selection, Morphology, Support vector machines

1 引言

如何准确地识别图像信息,理解图像的包含的模式,一直是计算机视觉的重要研究领域。遥感卫星图像数据近年来在地质勘探、农业监测、卫星导航、气象分析等领域得到了广泛应用,如何利用计算机快速并准确地识别并提取其中的地物类别信息,已经成为了计算机视觉和模式识别的一个重要研究方向^[1,2]。

遥感卫星图像数据可以记录丰富的地表信息,特别是高光谱传感器,如 HYDICE、ROSIS 等机载高光谱传感器平台,可以提供高达 5nm 的光谱分辨率,1~5m 空间分辨率的地面图像^[3]。由于高光谱遥感图像光谱分辨率很高,其表征和区分不同地物能力很强。但是,其波段数目巨大、波段间相关性高,处理高光谱图像数据所需的存储能力、计算时间、复杂度会很高。因此,对高光谱图像进行分类前,必须对数据进行特

征选择,以提取图像的主要特征。

特征选择在遥感图像数据中也称波段选择,主要是为了压缩波段数,同时尽可能地保持有用的判别信息,通过从原始高维数据中选择特征子集,达到保持最佳类别可分性的同时又不改变原始图像的光谱特征,也更有利于提取形态学特征。通过选取最佳的波段子集,可以不丢失任何重要的信息,且保持其物理特性。传统的波段选择方法包括基于 PCA 的波段选择^[4,5]、基于信息熵的波段选择^[6,7]和波段分组^[8]等方法。波段分组方法通过选取最佳波段子集,也取得了很好的应用。本文重点研究利用波段分组方法,同时考虑波段间的相关性和类别间的可分性来选择波段,提高地物的可分性。

在特征选择的基础上,传统方法一般将光谱特征输入到分类器中进行分类,缺少对空间信息的考虑。近年来有研究通过提取图像的纹理特征或形态学特征克服传统分类方法缺少空间信息这一缺点^[9]。纹理特征主要是评价一幅图像灰度

到稿日期:2014-02-07 返修日期:2014-04-21 本文受国家自然科学基金项目(61102128),国家自然科学基金重点项目(41061130553),国家重点基础研究发展计划(2012CB719905,2011CB707105)资助。

张帆(1990-),男,博士生,主要研究方向为图像场景解译和高光谱遥感影像处理;杜博(1983-),男,副教授,主要研究方向为模式识别、图像处理;张良培(1962-),男,教授,博士生导师,主要研究方向为遥感科学与技术、人工智能、模式识别等;张乐飞(1986-),男,博士后,主要研究方向为机器学习和图像处理。

的变化。比较典型的图像纹理特征有灰度共生矩阵 (GLCM)^[10,11]、Gabor^[12,13]。利用数学形态学来提取目标的形状特征是另一种比较常见的空间信息提取方法。在遥感图像中,利用形态学特征进行分类,已经被证实是一种非常有效的方法^[4,15]。不同尺度和形状的结构元素导致形态学特征维数会非常高,直接在原始高光谱图像提取,会导致提取后特征维数呈海量增加,分类出现明显的 HUGES 效应^[16]。因此,利用上述波段选择方法来选择最佳波段组合,再进一步提取形态学特征,是更加可行的方法。

在特征选择后提取图像的形态学特征,仍然会使分类时输入特征维数有较大增加。为了控制计算复杂度和时间,保持分类精度的稳定性,需要利用一定的特征提取算法来处理分类时输入形态学特征。递归特征降维算法 Recursive Feature Elimination (RFE) 是一种基于 SVM 的降维算法,可以按照特征不同维度的重要性对特征分量进行排列,并以 SVM 判决边界的变化作为准则来选取特征分量^[17],提取其中最重要的分量组成特征向量来进行分类。引入该算法,可以减少输入数据的维数。

本文提出了一种结合空间和光谱特征的高光谱图像分类方法。利用波段分组^[6]提取高光谱图像中的最佳波段子集,尽可能地保留光谱信息,减少波段个数;为了考虑不同大小的地物分布的空间特征,利用一定范围大小的结构元素来提取形态学特征,实现空间、光谱特征融合分类。

本文第 2 节介绍了波段分组特征选择算法;第 3 节介绍了形态学的基本原理;第 4 节介绍了 SVM 和 RFE 降维算法;第 5 节介绍了实验方法和结果;最后讨论并总结全文。

2 波段分组特征选择

对于高光谱图像^[3],由于其波段数较高,且波段范围窄,相邻波段具有很高的相关性。为了克服这一缺点,可以把光谱曲线等分为一个个较小的区域,在每一个区域中选择具有代表性的特征,如图 1 所示。该方法依赖于划分区域个数和区域边界的选择。为了更好地划分区域,也可以选择相邻特征相关性剧烈变化的边界作为划分边界。这样所得到的区域就不是等划分的,而且从这些区域中所选的特征相关性也会更小,如图 2 所示。

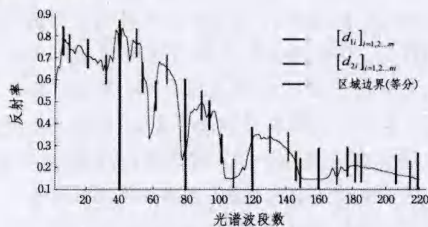


图 1 波段等分分组的特征选择

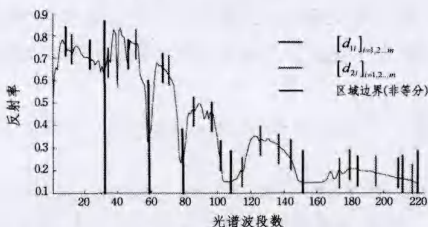


图 2 波段非等分分组的特征选择

为了得到最佳的划分边界,提出了一种波段分组算法,把波段分为连续但不同的波段组合。首先,由第一个波段开始,依次增加一个波段,直到所使用的判决阈值发生剧烈变化,则分组结束并开始下一个分组,最后得到许多组连续的波段分组。

假设 M_{i-1} 为未加入第 i 个波段的波段子集, M_i 为加入了第 i 个波段的波段子集,则波段分组的判决准则定义如下:

$$t = \frac{M_i - M_{i-1}}{M_{i-1}}$$

M_i 定义如下:

$$M_i = Bh_i \cdot Cor_i$$

当阈值 t 发生剧烈变化时,这里设置为 5%,则波段分组结束,开始新的波段子集分组。其中 M_i 为 Bhattacharya 距离和波段相关性的乘积, Bh_i 为波段子集间的 Bhattacharya 距离, Cor_i 为增加的 i 个波段和前面波段的最小波段相关性。

利用该方法,波段被分为 m 个不同但连续的波段分组,每一个分组中的波段都有很高的相关性,而分组间的波段相关性很小,这样我们可以提取每一个分组间最具有代表性的波段组成最佳波段子集,不仅可以最大限度地保留原始信息,还大大减少了数据量。然后进一步提取形态学特征。

3 数学形态学

图像的形态学是一个分析像素间空间关系的理论。对于高光谱图像,有许多种形态学算子可以用来提取几何特征。腐蚀和膨胀是形态学的两个基本操作。这些操作和不同的结构元素 (Structuring Elements) 一起来对图像进行处理。本文主要使用了以下的形态学算子。

3.1 开运算和闭运算

开运算和闭运算是腐蚀和膨胀的联合操作。开运算是针对一幅腐蚀后的图像进行膨胀处理,可以去掉图像中较亮的结构。相对地,闭运算是针对一幅膨胀后的图像进行腐蚀处理,可以去掉图像中较暗的结构。图 3 中展示了一系列不同尺度的开(闭)运算,可以看出图像上某些消失的结构。

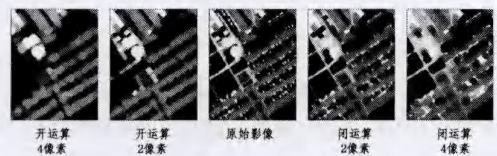


图 3 不同尺度的开(闭)运算,使用了圆形结构元素

3.2 开重构和闭重构

重构算法是一种迭代计算的形态学算子,由原始图像和掩膜构成。如果掩膜是原始图像的腐蚀结果,那么经过重构运算会把原始图像上较亮的结构去除。相对地,如果掩膜是原始图像的膨胀结果,那么经过重构运算会把原始图像上较暗的结构去除。图 4 展示了重构算法的结果。

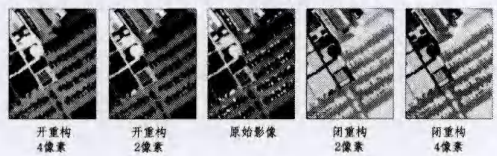


图 4 不同尺度的开(闭)重构,使用了圆形结构元素

4 SVM 分类和 RFE 特征选择

4.1 SVM

相较于其他非参数和统计方法,SVM 有着更高的分类精

度,而且在处理高维数据分类上表现出了很强的稳定性。SVM也是一种最常用的核学习算法,它寻找特征空间中的最佳分割超平面来实现分类。文献[18]对SVM分类的原理进行了详细的介绍。

支持向量机作为一种非参数化的方法,在进行求解前需要确定一些参数,对于高斯核,有两个参数。常用交叉验证的方法来寻找最优参数。本文中利用了LIBSVM工具包来求解^[19]。

4.2 RFE 特征选择

对于不同的图像数据,不同的特征可能有着不同的效果,形态学作为一种提取不同尺度大小目标和结构信息的算法,只能人为设定提取一定范围一系列尺度的特征,由于会形成大量的特征,因此在分类中无法确定其中哪些尺度的特征是最有效的。本文使用了递归特征降维算法RFE,来提取对分类最有效的特征,以分析不同尺度的形态学特征对于分类结果的影响。利用这种算法,不仅可以大大减少提取出来的形态学特征的数量,还可以确定哪些尺度的特征对分类影响最大。该方法以SVM判决边界的变化作为准则,来选取特征^[17]。

对于SVM的决策函数,如式(1)所示:

$$f(x) = w \cdot x + b \quad (1)$$

这里主要考虑非线性可分的情况,原始空间 x 通过映射,投影到一个更高维的空间。利用核函数计算特征空间的点积。通过计算权向量 w^2 ,如式(2)所示,就可以计算出每一个特征对决策函数的贡献。

$$W^2(\alpha) = \|w\|^2 = \sum_{k,l} \alpha_k \alpha_l y_k y_l k(x_k, x_l) \quad (2)$$

通过这一性质,就可以计算出对边界作用最小的特征,然后移除这一特征,达到降维的效果。如式(3)所示,使公式变化最小的特征将被移除。

$$f = \arg \min_{i \in Q} |W^2(\alpha) - W^2_{(-i)}(\alpha)| \quad (3)$$

上式仅仅考虑了二分类问题,为了更好地考虑多类别的情况,可以分别计算每一类别的权向量,如式(4)所示,选择决策变为计算所有类别的权向量之和。

$$f = \arg \min_{i \in Q} |W_a^2(\alpha) - W_{a,(-i)}^2(\alpha)| \quad (4)$$

RFE算法通过迭代计算,每一次迭代移除一个特征。最后,不仅可以达到降维效果,还可根据每一特征的贡献将其依次由大到小排列。实验流程如图5所示。

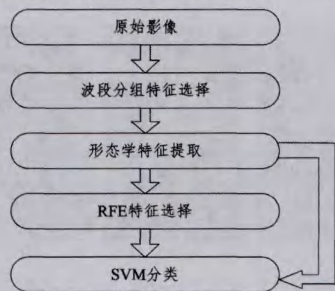


图5 实验流程

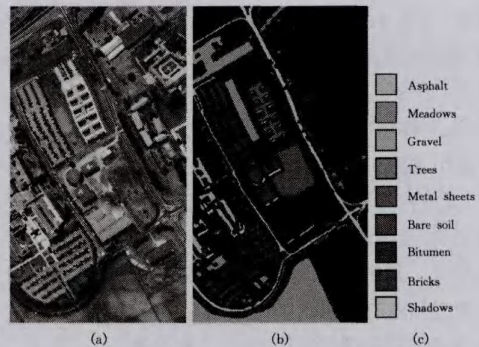
5 实验方法和结果

5.1 实验数据

本次实验采用了两组ROSIS的城区高光谱高空间分辨率图像。数据拍摄于意大利pavia市,由机载ROSIS传感器获得。

第一组数据覆盖了pavia大学区域,样本数目大小为610*

340,波段数为103个,一共包含9种地物类别。地物真实分布如图6所示,训练样本从每一类地物真实分布中随机选取50个作为训练样本,具体见表1。



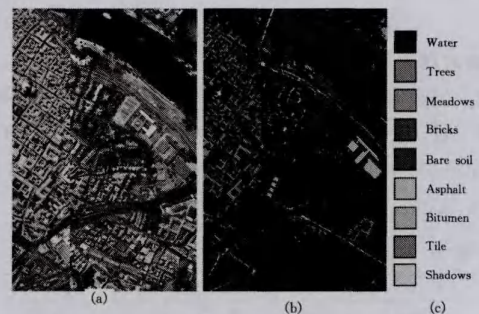
(a)是实验数据的假彩色合成图;(b)是地面地物的真实分布;
(c)是9种真实地物及其颜色

图6 pavia大学

表1 pavia大学实验数据测试和训练样本

No	Name	Samples	
		Train	Test
1	Asphalt	50	9248
2	Meadows	50	3090
3	water	50	65971
4	Trees	50	7598
5	Tile	50	42826
6	Soil	50	6584
7	Bitumen	50	7287
8	Bricks	50	2685
9	Shadows	50	2863
Total		450	148152

第二组数据覆盖了pavia市区,样本数目大小为1096*715,波段数为102个,一共包含9种地物类别。地物真实分布如图7所示,训练样本从每一类地物真实分布中随机选取80个作为训练样本,具体见表2。



(a)是实验数据的假彩色合成图;(b)是地面地物的真实分布;
(c)是9种真实地物及其颜色

图7 pavia市区

表2 实验数据测试和训练样本

No	Name	Samples	
		Train	Test
1	Asphalt	80	6641
2	Meadows	80	18649
3	Gravel	80	2099
4	Trees	80	3064
5	Metal Sheet	80	1345
6	Soil	80	5029
7	Bitumen	80	1330
8	Bricks	80	3862
9	Shadows	80	947
Total		720	42776

5.2 实验步骤

实验中波段分组特征使用的判决矩阵是波段相关性和 Bhattacharya 距离的乘积,每一分组包含的最大波段数是 18,最小是 8。实验的分类器基于 SVM,采用 RBF 核进行分类。RBF 核的参数选择利用了网格法交叉验证。C 和 G 的范围为 $[2^{-8}, 2^{-7}, \dots, 2^7, 2^8]$ 。对于每一组结果,分别计算了分类的混淆矩阵、总体精度和 KAPPA 系数。

实验中采用了 4 种形态学特征(见表 3),分别为开运算、闭运算、开重构、闭重构。对于每一组特征,采用了圆形的结构元素,半径范围为 2~10,步长为 2,每一种特征可产生 5 幅形态学图像。其中 Orin 代表原始数据,ALL 代表波段分组选择后数据和所有提取出来的形态学特征,RFE 是利用 ALL 数据进行降维后得到的特征,具体流程见图 8。

表 3 实验数据组成(Orin=原始图像,BG=波段分组选择后数据, O=开运算,C=闭运算,OR=开重构,CR=闭重构)

Data	Orin	BG	O	C	OR	CR
Orin	X					
ALL		X	X	X	X	X
RFE		X	X	X	X	X

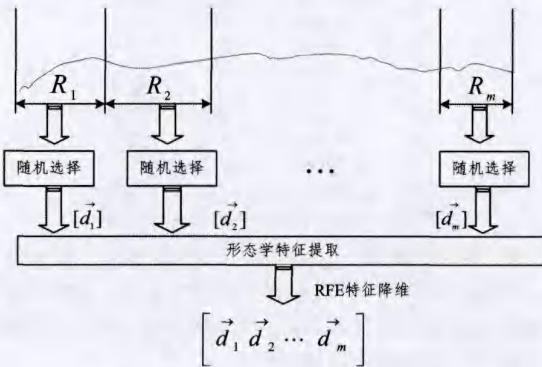


图 8 特征选择流程

5.3 实验结果与分析

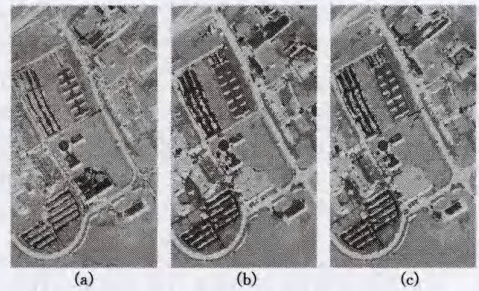
第一组数据实验结果如表 4 所列,分类结果如图 9 所示。从结果上看,仅仅利用原始图像并不能得到非常好的分类结果,大部分类别的分类精度都不够高,总体分类精度为 87.49%,Kappa 系数为 83.83。

表 4 实验数据 RFE 降维后分类精度和 Kappa 系数

Class	Orin	ALL	RFE11	RFE21
Asphalt	83.20	96.15	91.99	95.52
Meadows	86.17	94.90	91.75	92.59
Gravel	86.42	99.29	92.42	96.47
Trees	95.59	95.20	93.11	92.75
Metal-Sheet	99.85	100.00	99.93	99.93
Soil	90.79	97.20	98.13	98.63
Bitumen	93.83	96.77	96.47	96.54
Bricks	81.18	98.64	94.43	96.77
Shadows	100.00	99.37	100.00	99.89
Overall Accuracy	87.49	96.24	93.49	94.83
Kappa	83.83	95.07	91.49	93.24

对原始图像进行波段选择,分为 7 组,每组随机选择,构成 7 个波段,在提取每一波段形态学特征后,可以看到实验的总体分类精度有大幅度的提升。大部分地物类别的分类精度也都有明显的提高。结合形态学特征的分类精度达到了 96.24%,Kappa 系数达到了 95.07。在 Meadows 的分类精度

达到 94.90%,是所有实验中分类精度最高的一组。



(a) 是利用原始图像的结果;(b) 为 OC-OCR 数据集结果;
(c) 为 RFE-21 结果

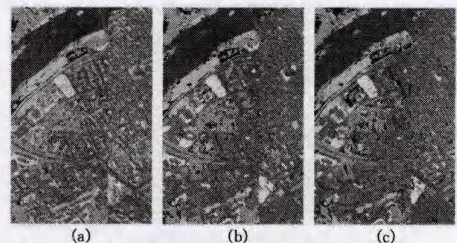
图 9 pavia 大学分类结果

为了考虑不同特征维数对分类结果的影响,并分析不同尺度形态学特征对结果的影响,一共运行了 2 组不同的 RFE 特征选择实验,分别选取了 11、21 个特征来进行分类。可以看出:1) 选择 11 个特征进行分类时,分类精度达到了 93.49%,只利用了 7.5%的原始特征,仅损失了 2.75%的分类精度,在 Shadows 精度上更是达到了 100%;2) 在选择 21 个特征进行分类时,精度有了一定的提升,在 Soil 类别上的精度为所有实验中最好的。

第二组数据实验对原始图像进行波段选择,分为 8 组,每组随机选择,构成 8 个波段,并提取每一波段形态学特征,结果如表 5 所列,分类结果如图 10 所示。从结果上看出,第二组数据在结合了形态学特征后,分类精度也得到了显著提高。

表 5 实验数据 RFE 降维后分类精度和 Kappa 系数

Class	Orin	ALL	RFE11	RFE21
Water	99.13	98.42	99.72	99.61
Tree	90.92	92.29	94.68	95.35
Meadow	93.79	95.31	96.05	96.86
Brick	90.84	99.29	98.14	98.99
Soil	90.81	97.83	99.50	99.65
Asphalt	95.95	97.13	95.73	95.86
Bitumen	89.39	94.77	97.46	97.85
Tile	98.12	99.29	98.72	98.77
Shadows	99.97	100.0	100.0	100.0
Overall Accuracy	97.12	98.05	98.70	98.77
Kappa	95.94	97.25	98.17	98.26



(a) 是利用原始图像的结果;(b) 为 OC-OCR 数据集结果;
(c) 为 RFE-21 结果

图 10 pavia 市区分类结果

第二组也运行了 2 组不同的 RFE 特征选择实验。可以看出:1) 利用 RFE 特征选择 11 个特征时,不仅大大降低了数据维数,总体分类精度还有了小幅度提高;2) 当选择 21 个特征,分类精度到达了最高,总体精度达到了 98.77%。

从以上两组数据中可以看出,RFE 特征选择算法不仅大幅度降低了分类时输入的数据特征维数,而且还很好地保持

了分类精度,在第二组实验中,更进一步提高了分类精度;而且利用 RFE 提取出来的特征,还可以确定不同尺度的形态学特征对于数据分类的重要性,以提取更有效尺度的形态学特征。

表 6 中列出了 RFE 提取出来的前 11 个最重要的特征,在第一组实验数据中,开闭重构对于数据的分类有着更显著的影响,而且所使用的结构元素的大小都集中在 8 和 10,对于这组数据,大尺度的形态学特征更有利于提高分类精度。在第二组实验数据中,开闭重构所提取的特征比例较大,而结构元素大小主要集中于 4 和 6,说明中等尺度的结构元素更适合于这组数据,也充分说明了对于不同的数据,不同尺度的结构对于分类的影响是不同的,利用合适尺度的形态学特征,可以更好地提高分类精度。

表 6 RFE 降维提取前 11 个特征(Band-#代表特征来自第#个波段分组,OP=开运算,CL=闭运算,OR=开重构,CR=闭重构,OP-#代表尺度为#的开运算)

Num	Pavia 大学	Pavia 市区
1	Band-6, OP-10	Band-2, OP-4
2	Band-5, CR-10	Band-3, CR-4
3	Band-3, CR-10	Band-6, CR-10
4	Band-7, OR-8	Band-4, CL-4
5	Band-7, OR-10	Band-5, CR-6
6	Band-3, OR-8	Band-8, CR-10
7	Band-1, CL-10	Band-7, CR-4
8	Band-4, OP-10	Band-6, OR-6
9	Band-2, CR-10	Band-4, CL-6
10	Band-3, CR-8	Band-8, OR-8
11	Band-6, OP-8	Band-3, CR-6

结束语 本文充分利用了高光谱图像丰富的光谱信息,利用波段分组算法提取图像中最具代表的特征来进行分类,避免了高维数据带来的数据冗余和计算复杂度的增加,最大限度地保留了高光谱的光谱信息,同时利用 RFE 算法提取出一定数量的形态学特征辅助光谱特征进行分类。

实验证明:1)高光谱图像拥有丰富的光谱信息,利用波段分组算法进行波段提取,能降低数据冗余,并结合一定的形态学特征是非常有效的,可以最大限度地保留判别性的光谱信息,并加入其他特征。2)利用 RFE 特征提取算法来对数据进行降维,选择必要尺度的形态学特征不仅可以大量减少输入数据的维数,而且还可以保持非常高的分类精度,极大提升分类的运行速度,减少计算时间。

未来的工作将进一步研究加入更多的特征来辅助分类和更有效的数据降维方法来应对更多的特征带来的维数急剧增加的问题。

参 考 文 献

[1] 钟燕飞,张良培,李平湘.基于多值免疫网络的多光谱遥感影像分类[J].计算机学报,2007,30(12):2181-2188

[2] 王兆虎,刘芳,焦李成.一种基于视觉特性的遥感图像分割[J].计算机学报,2005,28(10):1686-1691

[3] Goetz A F H, Vane G, Solomon J E, et al. Imaging spectrometry for earth remote sensing[J]. Science, 1985(228): 1147-1153

[4] Hsu Pai-hui. Feature extraction for hyperspectral image [C]//

Proceeding of the 20th Asian Conference Remote Sensing. 1999: 408-410

[5] Yang Zhu-sheng, Guo Lei, Luo Xin. Research on segmented PCA based on band selection algorithm of hyperspectral image [J]. Engineering of Surveying and Mapping, 2006, 15: 15-18

[6] Chavez P S, Berlin G L, Sowers L B. Statistical method for selecting Landsat MSS Ratios [J]. Journal of Applied Photogrammetric Engineering, 1982, 1: 23-30

[7] Samiappan S, Prasad S, Bruce L M. Non-Uniform random feature selection and kernel density scoring with svm-based ensemble classification for hyperspectral image analysis [J]. IEEE Journal of PP Selected Topics in Applied Earth Observations and Remote Sensing, 2013(99): 1-9

[8] Prasad S, Bruce M. Decision fusion with confidence-based-weight assignment for hyperspectral target recognition [J]. IEEE Transactions Geoscience and Remote Sensing, 2008, 46: 1448-1456

[9] Soille P, Pesaresi M. Advances in mathematical morphology applied to geoscience and remote sensing [J]. IEEE Transactions Geoscience and Remote Sensing, 2002, 40(9): 2042-2055

[10] Haralick R M, Shanmugam K, Dinstein I. Textural features for image classification [J]. IEEE Transactions on Systems, Man, and Cybernetics, 1973, SMC-3(6): 610-621

[11] Fauvel M, Tarabalka Y, Benediktsson J A, et al. Advances in Spectral-Spatial classification of hyperspectral Images [J]. Proceedings of the IEEE PP, 2012(99): 1-24

[12] Kruizinga P, Petkov N, Grigorescu S E. Comparison of texture features based on Gabor filters [C]// Proc. Int. Conf. Image Anal. Process. . 1999: 142-147

[13] Clausi D A, Deng H. Design-based texture feature fusion using Gabor filters and co-occurrence probabilities [J]. IEEE Transactions Image Processing, 2005, 14(7): 925-936

[14] Pesaresi M J, Benediktsson A. A new approach for the morphological segmentation of high-resolution satellite images [J]. IEEE Transactions Geoscience and Remote Sensing, 2001, 39(2): 309-320

[15] Fauvel M, Benediktsson J A, Chanussot J, et al. Spectral and spatial classification of hyperspectral data using svms and morphological profiles [J]. IEEE Transactions Geoscience and Remote Sensing, 2008, 46(11): 3804-3814

[16] Hughes G. On the mean accuracy of statistical pattern recognizers [J]. IEEE Transactions on Information Theory, 1968, 14(1): 55-63

[17] Tuia D, Pacifici F, Kanevski M, et al. Classification of very high spatial resolution imagery using mathematical morphology and support vector machines [J]. Geoscience and remote sensing, IEEE Transactions, 2009, 47(11): 3866-3879

[18] Burges C J C. A tutorial on support vector machines for pattern recognition [J]. Data Mining and Knowledge Discovery, 1998, 2: 212-167

[19] Chang C C, Lin C J. LIBSVM: A library for support vector machines [OL]. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>