

文章编号:1001-9014(2012)02-0166-06

基于最大信息量的高光谱遥感图像 无监督波段选择方法

刘雪松¹, 葛亮¹, 王斌^{1,2}, 张立明¹

(1. 复旦大学 电子工程系, 上海 200433; 2. 复旦大学 波散射与遥感信息重点实验室, 上海 200433)

摘要:提出一种基于最大信息量的高光谱遥感图像无监督波段选择方法. 该方法以在所选择的波段中保留原数据中所包含的最大信息量为目标, 并采用逐个移除波段的方式来实现. 算法使用 K-L 散度来定量表示信息量的大小, 并通过信息量在整个数据集中的分布情况来决定所移除的波段. 与传统方法相比, 具有物理意义明确、计算过程简单的优点, 同时还能够完全自动地完成, 实现无监督的波段选择.

关键词:高光谱遥感图像; 波段选择; 信息量; K-L 散度; 分类

中图分类号:TP751 **文献标识码:**A

An unsupervised band selection algorithm for hyperspectral imagery based on maximal information

LIU Xue-Song¹, GE Liang¹, WANG Bin^{1,2}, ZHANG Li-Ming¹

(1. Department of Electronic Engineering, Fudan University, Shanghai 200433, China;

2. The Key Laboratory of Wave Scattering and Remote Sensing Information, Fudan University, Shanghai 200433, China)

Abstract: An unsupervised band selection algorithm for hyperspectral imagery based on maximal information is proposed in this paper. The objective of the method is to preserve the maximal information from original data in the selected bands. The bands with less information are removed one by one from the original data. K-L divergence is used to quantify the information amount and its distribution over all the dataset is considered to judge the specific band which needs to be removed. Compared with traditional methods, the proposed approach has an explicit physical meaning and its computational process is very simple. It is an unsupervised method and can perform automatically.

Key words: hyperspectral imagery; band selection; information amount; K-L divergence; classification

PACS: 84.40.X6

引言

高光谱遥感的主要特点是成像光谱仪在电磁波谱的紫外、可见光、近红外和中红外区域中的数十至数百个非常窄且连续的光谱段上同时获取信息, 得到每个像素的完整连续的光谱曲线^[1]. 获得较高光谱分辨能力的同时, 也带来了数据量的增大. 且高光谱图像波段之间的相关系数往往很高, 因而包含了大量的冗余信息, 会造成存储和处理能力的浪费, 如何从原始数据中剔除这些冗余信息, 在不损失重要信息的情况下尽可能降低数据的维度, 成为一个重要问题.

针对这个问题, 可以采用波段选择的方法, 从原始的数百个波段中选出最能表示整体特征的数十个或数个波段, 并以此来表示整个数据集^[2]. 当前对高光谱图像波段选择方法的研究, 都是针对其在分类中的应用, 具体又可以分为有监督和无监督两种^[3]. 有监督的方法通常需要手动输入训练样本, 使算法符合实际数据的特点, 以达到波段选择后实现较准确分类的目的^[4]. 无监督的方法不使用任何先验知识, 这样就能实现完全自动的波段选择. 本文的研究主要针对无监督的方法.

当前对无监督波段选择方法的研究大部分都是

收稿日期: 2011-05-06, 修回日期: 2011-12-12

Received date: 2011-05-06, revised date: 2011-12-12

基金项目: 国家自然科学基金(61071134); 863 国家高技术研究计划(2009AA12Z115)

Foundation items: Supported by National Natural Science Foundation of China (No. 61071134) and National High Technology R&D Program (No. 2009AA12Z115)

作者简介: 刘雪松(1987-), 男, 江苏扬州人, 硕士研究生, 主要从事遥感图像处理的研究. E-mail: 082021037@fudan.edu.cn.

基于波段排序的思想,如 ID (Information Divergence)^[5]、MVPCA (Maximum-Variance Principle Component Analysis)^[6]等.也有一些基于聚类的方法,如 AP(Affinity Propagation)^[7]等,AP方法本是一种在计算机视觉中应用广泛的聚类方法^[8],近年来也被应用到波段选择当中.本文提出了一种基于最大信息量的无监督波段选择方法.对高光谱图像的后续处理通常都是一个从原始信息中挖掘其规律的过程,因此以被选择的波段中含有最大的信息量为算法的指标,有着很明确的物理意义.为了能够定量地表示信息量的大小,使用了 K-L 散度的概念.通过分析信息量在整个数据集中的分布情况,采用迭代的方法,逐个移除对数据集的信息量影响最小的波段,直到达到最后的目标.

1 算法描述

1.1 K-L 散度与信息量分布

直接衡量高光谱图像每个波段中所包含的信息量大小,是比较困难的.由于波段选择的主要目的是为了去除邻近波段之间大量的冗余信息,因此可以对波段之间的信息量差进行分析,并以此为根据选择和移除冗余的波段.这里使用 K-L 散度来定量衡量这种信息量差.

K-L 散度是一种在信息论中得到广泛应用的信号相似度量^[9].若两个离散随机信号的概率分布函数分别为 $\mathbf{P} = [P_1, \dots, P_n, \dots, P_N]^T$ 和 $\mathbf{Q} = [Q_1, \dots, Q_n, \dots, Q_N]^T$,并且有 $\sum_{n=1}^N P_n = \sum_{n=1}^N Q_n = 1$. 那么定义 \mathbf{Q} 相对于 \mathbf{P} 的 K-L 散度为

$$D_{KL}(\mathbf{P} \parallel \mathbf{Q}) = \sum_{n=1}^N P_n \log \frac{P_n}{Q_n}. \quad (1)$$

式(1)的物理意义是用 \mathbf{Q} 中的元素来表示 \mathbf{P} 中的所有元素所需要的额外信息量的大小.如果这里的 \log 是以 2 为底的对数,结果单位为 bit;如果 \log 是自然对数,结果单位为 nat. K-L 散度越大,说明完全用 \mathbf{Q} 中的元素来表示 \mathbf{P} 的难度越大,也即两个信号之间的相似度越低.因此, K-L 散度本质上是一种非对称的概率意义上的距离,或者可以理解为两个信号所包含的信息量之差^[10].

将 K-L 散度应用到两个波段的图像之间,就可以得到波段之间的信息量差异.假设一组高光谱数据 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_L]^T \in R^{L \times N}$,其中 L 和 N 分别表示图像的波段个数和像素个数, $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{iN}]^T$ 表示第 i 个波段.再假设 $\tilde{\mathbf{x}}_i = \mathbf{x}_i / \sum_{n=1}^N x_{in}$ 为归一化后的 \mathbf{x}_i ,

那么第 j 个波段相对于第 i 个波段的 K-L 散度为

$$D_{KL}(\tilde{\mathbf{x}}_i \parallel \tilde{\mathbf{x}}_j) = \sum_{n=1}^N \tilde{x}_{in} \log \frac{\tilde{x}_{in}}{\tilde{x}_{jn}}. \quad (2)$$

式(2)的物理意义是用第 j 个波段来表示第 i 个波段所需要的额外信息量的大小.考虑波段 i 和 j 所构成的集合,如果从中移除波段 i ,便损失了大小为 $D_{KL}(\tilde{\mathbf{x}}_i \parallel \tilde{\mathbf{x}}_j)$ 的信息量.反之如果移除波段 j ,则会损失 $D_{KL}(\tilde{\mathbf{x}}_j \parallel \tilde{\mathbf{x}}_i)$ 的信息量.

为了分析整个数据集中所有波段所包含信息量的分布情况,定义高光谱数据集 $\mathbf{X} \in R^{L \times N}$ 的“信息量分布表”为:

$$\mathbf{D} = \begin{bmatrix} 0 & D_{KL}(\tilde{\mathbf{x}}_1 \parallel \tilde{\mathbf{x}}_2) & \dots & D_{KL}(\tilde{\mathbf{x}}_1 \parallel \tilde{\mathbf{x}}_L) \\ D_{KL}(\tilde{\mathbf{x}}_2 \parallel \tilde{\mathbf{x}}_1) & 0 & & \vdots \\ \vdots & & \ddots & D_{KL}(\tilde{\mathbf{x}}_{L-1} \parallel \tilde{\mathbf{x}}_L) \\ D_{KL}(\tilde{\mathbf{x}}_L \parallel \tilde{\mathbf{x}}_1) & \dots & D_{KL}(\tilde{\mathbf{x}}_L \parallel \tilde{\mathbf{x}}_{L-1}) & 0 \end{bmatrix}, \quad (3)$$

也即

$$D(i, j) = D_{KL}(\tilde{\mathbf{x}}_i \parallel \tilde{\mathbf{x}}_j), \quad (4)$$

其中, $D(i, j)$ 表示 \mathbf{D} 的第 i 行第 j 列的元素.

信息量分布表反映了信息量在整个数据集内部分布的情况,可以进一步可视化为“信息量分布图”.实际的高光谱遥感图像的信息量分布见图 1.它们分别是由机载可见光及红外成像光谱仪(Airborne Visible/Infrared Imaging Spectrometer, AVIRIS)拍摄于美国印第安纳州 Pine 测试点的 Indiana 数据(<http://cobweb.ecn.purdue.edu/~biehl/Multi-Spec>)和美国内华达州 Cuprite 地区的数据(<http://aviris.jpl.nasa.gov/html/aviris.freedata.html>),以及由高光谱图像收集实验仪器(Hyperspectral Digital Imagery Collection Experiment, HYDICE)拍摄于美国德克萨斯州 Copperas Cove 镇的 Urban 数据(<http://www.tec.army.mil/Hypercube>).图中蓝色的部分表示波段之间的信息量差较小,而红色表示信息量差较大.因而在波谷的部分存在着大量的冗余信息.

1.2 通过迭代方法逐个移除波段

考虑 \mathbf{D} 中的任意一行 $\mathbf{d}_i = [D(i, 1), D(i, 2), \dots, D(i, L)]$,它反映的是用其他各波段来表示波段 i 所需要的额外信息量,也即波段 i 比其他各波段多出的信息量.假设 \mathbf{d}_i 中除了 $D(i, i) = 0$ 之外最小的元素是 $D(i, j)$,那么如果将波段 i 从整个数据集中移除,则波段 i 的一部分信息(大小为 $D(i, j)$)随着它被移除而丢失了,而另一部分信息依然包含在剩下的波段当中最接近 i 的波段,也即波段 j 中.因此

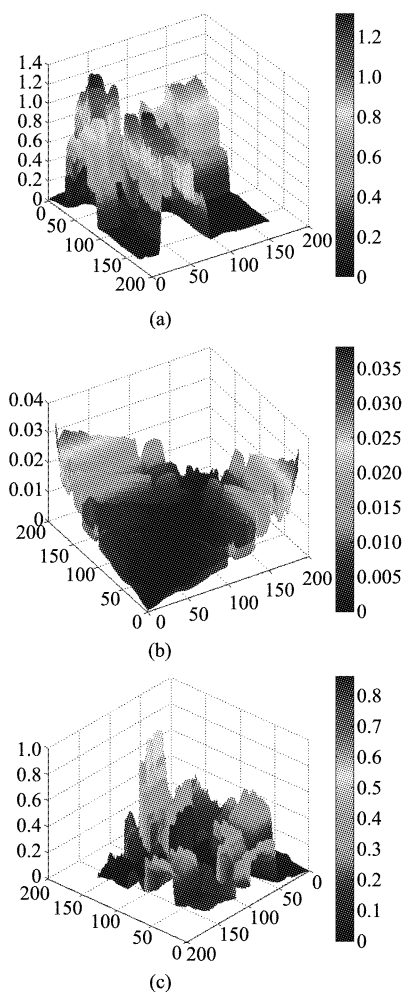


图1 几组实验数据的信息量分布图(a) Indiana 数据, (b) Cuprite 数据, (c) Urban 数据

Fig.1 The abundance distribution map for some experimental data sets (a) Indiana dataset, (b) Cuprite dataset, (c) Urban dataset

移除波段 i 对整个数据集所造成的信息量损失应该等于 $D(i, j)$. 也就是说, 波段 i 对整个数据集的信息量贡献为 $D(i, j)$.

图2用信息量分布图说明了这个概念, 沿图中的横坐标轴作一个对应着某波段的切面, 并考虑该切面上的值, 如图中的白色线条所示. 很明显, 该线条的波谷位置(箭头所指位置)的值反映了该波段对整个数据集的信息量贡献的大小.

为了方便表述, 将 \mathbf{D} 中的所有对角线元素修正为大数 C , 通常可以取 \mathbf{D} 中元素的最大值. 也即

$$D(i, j) = \begin{cases} D_{KL}(\tilde{\mathbf{x}}_i \| \tilde{\mathbf{x}}_j), & i \neq j \\ C, & i = j \end{cases}, \quad (5)$$

那么可以定义波段 i 对整个数据集的信息量贡献为 $\text{contribution}(i) = \min(\mathbf{d}_i)$, (6) 它同时也等于从数据集中移除波段 i 时所损失的信

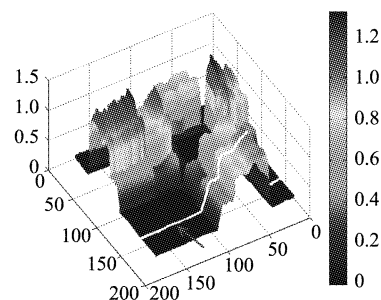


图2 波段移除的原理(其中底面上两个方向的坐标分别表示两个波段的序列号, 纵坐标表示它们之音质信息量差)
Fig.2 The principle of band removal (Coordinates in the sub-face is the indices of two bands, while the ordinate stands for the information difference between the two bands)

息量.

现在可以将最大信息量的问题转化为移除信息量贡献最小的波段. 在每次从数据集中移除一个波段之后, 将剩下的波段看作是一个新的数据集, 并重复使用以上方法, 直至达到预先设定的目标.

值得注意的是, 这里只能采用逐个移除的方式, 而不能同时移除多个波段. 这是因为所描述的信息量贡献只是一个相对的概念, 一个贡献很小的波段并不是因为它包含的信息量很少, 而是因为它所包含的信息量同时也包含在其他波段中, 因而信息量的损失也是一个相对的概念. 假设数据集中存在一组很相似的波段, 由于它们之间包含的信息量大多是重复的, 因此它们的信息量贡献都很小. 如果采用同时移除多个波段的方法, 这一组波段会全部被移除, 从而造成大量的信息量损失. 而若采用逐个移除的方式, 当该组波段被移除到只剩一个时, 由于与其相似的波段已经都被移除, 它的信息量贡献值会变成一个较大的值, 从而被保留下来.

算法的具体流程如下:

算法 基于最大信息量的波段选择方法 (Maximal Information, MI)

给定观察矩阵 $\mathbf{X} \in R^{L \times N}$ 和需要选择的波段数 K .

步骤1 根据式(2)和式(5)计算 $\mathbf{D}(i, j)$, 其中 $C = \max(\mathbf{D}(i, j))$. 设置初始标号集 $\text{IND} = \emptyset$.

步骤2 根据式(6)寻找 $G = \arg \min_i (\text{contribution}(i))$, 并令 $\text{IND} = \text{IND} \cup G$.

步骤3 令 $D(G, i) = D(i, G) = C, (i = 1, 2, \dots, L)$.

步骤4 如果 IND 中元素的个数小于 $L - K$, 回到步骤2.

步骤5 根据 IND 中的标号, 从 \mathbf{X} 中移除对应

的行,并输出结果.

以上算法中的步骤 3 将矩阵 D 的第 i 行和第 i 列分别全部替换为大数 C ,其作用相当于从整个信息量分布表中移除了第 i 个波段,并将移除后的数据集作为一个新的数据集进行下一次迭代.

2 实验验证

本文所提出的方法和前文所提到的几种非监督的方法 MVPCA^[6]、ID^[5] 和 AP^[7] 进行了比较.通过实验来测试几种波段选择算法在应用到高光谱图像分类时的效果.实验中所使用的分类方法有最近邻(KNN, K-nearest Neighbourhood)^[11],线性判别式分析(LDA, Linear Discriminant Analysis)^[12],径向基函数网络(RBF, Radial Basis Function)^[13]和支撑向量机(SVM, Support Vector Machine)^[14].

实验所采用的数据集是前文所提到的 Indiana 数据.它成像于 1992 年 6 月,成像区域为美国印第安那州的派恩遥感测试点,该数据有 220 个波段,波长范围从 0.4 ~ 2.5 μm ,光谱分辨率为 10 nm,空间分辨率为 17 m.实验所用的图像大小为 145 \times 145.该数据已被广泛地用于遥感图像的分类研究. Purdue 大学已经给出一份关于该区域的实地调查报

告^[15].该地区是位于印第安纳州西部城市西拉斐特(West Lafayette, IN)西北方向约 10 km 处的一片农田,覆盖该区域的主要是各种农作物(大约占三分之二,包括玉米、小麦、大豆、干草堆)和天然植被(大约占三分之一,由树林、草地等组成).除此之外,还有一些人工用地,包括区域顶部的一条双向高速公路(U. S. 52 和 U. S. 231)和一条铁路、中间的一条 2 级公路(杰克逊高速公路)、区域上方的一个无线电发射塔以及一些零星的房屋.取第 70、86、136 波段分别作为 R、G、B 分量合成伪彩色图,如图 3(a)所示.在进行处理之前,该数据的第 1 ~ 4, 78 ~ 82, 103 ~ 115, 148 ~ 166 以及 211 ~ 220 波段由于信噪比太低或为水吸收波段而被移除,剩下 169 个波段被用于进一步处理.地物真实中共给出了 16 类不同的地物,但是其中有 7 类的样本数过少,不具有代表性^[7],因而我们不考虑这些地物,从而设置分类个数为 9,其对应的地物真实如图 3(b)所示.这 9 个类各自所对应的地物和其分类样本个数如表 1 所示.

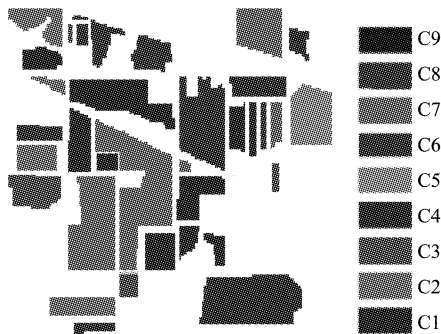
表 1 Indiana 数据中的地物分类样本个数

Table 1 Number of samples for ground objects in Indiana dataset

类别	对应地物	分类样本个数
C1	玉米未耕地(Corn-notill)	1256
C2	玉米疏耕地(Corn-min)	726
C3	牧草(Grass/Pasture)	431
C4	林地(Grass/Tress)	626
C5	干草(Hay-windrowed)	443
C6	大豆未耕地(Soybeans-notill)	828
C7	大豆疏耕地(Soybeans-min)	2284
C8	大豆已耕地(Soybeans-clean)	503
C9	树木(Woods)	1198



(a)



(b)

图 3 Indiana 数据(a)伪彩色图,(b)地物真实
Fig. 3 Indiana dataset (a) False-color image, (b) Ground truth

首先对波段选择后的信息量进行分析.设置波段个数为 10 个,分别用几种方法对原始数据进行波段选择.对波段选择后的数据分别计算它们的信息量分布表,然后得到每一个被选出的波段的信息量贡献.表 2 中给出了各种方法所得到的结果的波段序列号,以及它们对整体的贡献值之和.从表 2 可以看出,MI 方法得到的结果的信息量贡献最大,也即包含了最多的信息量.

表 2 各方法波段选择结果的信息量贡献对比

Table 2 Comparison of the information distribution among different band selection methods

波段选择方法	MI	MVPCA	AP	ID
波段选择结果	1, 25, 31, 32, 33, 72, 88, 92, 126, 129	17, 31, 32, 33, 67, 81, 90, 93, 126	25, 31, 32, 38, 40, 76, 89, 90, 91, 126	12, 31, 32, 33, 34, 73, 86, 90, 93, 113
信息量贡献	0.4808	0.3757	0.2547	0.3438

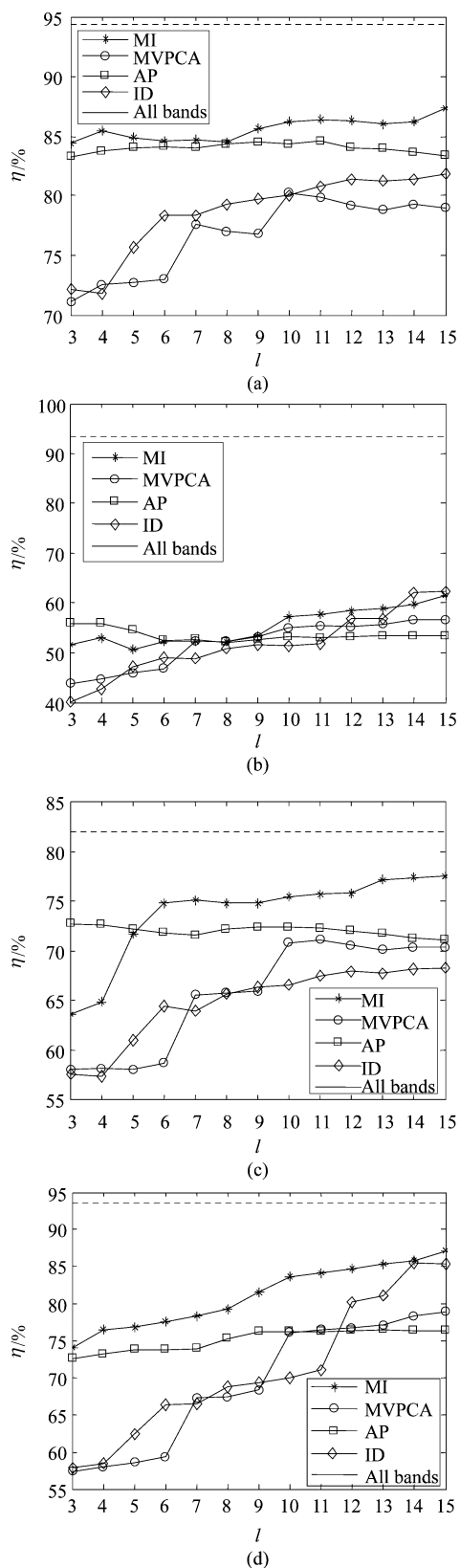


图4 分类精度与所选择波段个数的关系 (a) KNN 方法, (b) LDA 方法, (c) RBF 方法, (d) SVM 方法

Fig. 4 Relationship between the classification accuracy and the number of selected bands (a) KNN method, (b) LDA method, (c) RBF method, (d) SVM method

接下来分别设置波段个数为 3 ~ 15 个进行波段选择, 然后使用上述的分类方法对数据进行分类. 使用 KNN、LDA、RBF 和 SVM 分类的结果分别如图 4 所示. 其中横坐标均为波段个数, 而纵坐标均为分类精度. 图中的虚直线表示的是使用全波段所得到的分类精度. 从图中可以看出, 对于 KNN 而言, MI 和 MVPCA 的效果要好于另外两种方法, 而 MI 又要好于 MVPCA. 而在 LDA 的实验中, 所有方法的性能相差不是很大, 并且均处于一个较低的水平, 这是因为 LDA 方法本身需要降维, 因而其在本身分类精度不太理想的同时, 对不同的波段选择方法也不太敏感. RBF 的分类精度也不是很高, 其中当选取波段个数大于 6 个时, MI 的效果明显好于其他的方法. 而在 SVM 的实验中, MI 的效果要明显好于其他的方法. 总体而言, MI 方法的性能优于其他方法.

3 结论

提出了一种基于最大信息量的无监督波段选择方法. 高光谱图像波段之间的相关系数往往很高, 因而包含了大量的冗余信息, 在进行处理时这些冗余信息必然会造成运算能力的浪费. 设计了一种新的目标函数, 它以所选择的波段中包含的信息量最大为目标, 并通过迭代的算法, 每次移除一个使总信息量损失最小的波段, 直至达到预设的波段数目. 该方法的物理意义明确, 并且易于实施. 实验证实, 本文所提出的方法在应用到高光谱图像的分类时能够取得良好的效果, 并且性能优于其他已有的无监督方法.

将来的工作主要包括两个方面. 首先, 目前的方法必须要事先指定选择波段的数目, 未来我们希望能够使算法自动判断最佳的波段数. 此外, 目前的迭代算法虽然保证了每一步的最优解, 但是经多次迭代后最终得到的结果不一定是全局最优解, 如果改进的算法能够进行整体考虑, 直接寻找全局最优解, 或许可以更进一步提高其性能.

REFERENCES

- [1] Chang C-I. *Hyperspectral imaging: techniques for spectral detection and classification* [M]. New York: Plenum, 2003.
- [2] Guyon I, Elisseeff A. An introduction to variable and feature selection [J]. *J. Mach. Learn. Res.*, 2003, **3**: 1157 - 1182.
- [3] Keshava N. Distance metrics and band selection in hyperspectral processing with applications to material identification and spectral libraries [J]. *IEEE Trans. Geosci. Remote Sens.*, 2004, **42**(7): 1552 - 1565.

来增加数据量的处理能力。

4 结论

以满足配准机载前下视红外图像与可见光图像从而实现图像导航的实际需求为牵引,通过分析传感器成像的几何关系,提出了一种基于传感器参数和改良 CPD 算法的红外与可见光图像自动配准算法。首先,利用传感器的姿态参数和高度信息,对前下视红外图像进行几何透视校正,消除图像间的旋转和比例缩放等差异;然后,对可见光卫星图像和几何校正后的红外图像提取边缘特征点,基于相似变换模型,利用改良的 CPD 算法对其实现精配准。本文的主要贡献总结起来有两点:(1)提出了一种可有效解决前下视红外图像与可见光图像配准的系统方法;(2)提出了一种改良的 CPD 算法,避免了原始 CPD 算法对权重参数‘ w ’的人工选取,通过 PSO 优化算法完成了对权重参数‘ w ’的自动求解,从而优化了 CPD 算法的性能并扩展了其应用空间。最后,基于实测红外图像数据,验证了本文所提算法的有效性,实现配准误差精度约为 1 个像素,为其在图像导航领域的重要军事应用提供了一种理论思路。

REFERENCES

[1] LIU Jing, SUN Ji-Yin, ZHU Jun-Lin, *et al.* FLIR scene matching algorithm for complex ground target[J] *Application*

Research of Computers (刘婧,孙继银,朱俊林,等.复杂地面目标前视红外景象匹配算法.计算机应用研究), 2010, **27**(1):350-352.

- [2] YUAN Jin-Sha, ZHAO Zhen-Bing, GAO Qiang, *et al.* Review and prospect on infrared/visible image registration [J]. *Laser & Infrared* (苑津莎,赵振兵,高强,等.红外与可见光图像配准研究现状与展望.激光与红外), 2009, **39**(7):693-699.
- [3] Lee J H, Kim Y S, Lee D, *et al.* Robust CCD and IR image registration using gradient-based statistical information [J]. *IEEE Signal Processing Letters*, 2010, **17**(4):347-350.
- [4] Makrogiannis S, Bourbakis N G. Efficient registration of multitemporal and multisensor aerial images based on alignment of nonparametric edge features [J]. *Journal of Electronic Imaging*, 2010, **19**(1):1-15.
- [5] Hrkac T, Kalafati Z, Krapac J. Infrared-visual image registration based on corners and Hausdorff distance [J]. *Image Analysis*, 2007, **4522**:383-392.
- [6] Myronenko A, Song X. Point set registration: Coherent point drifts [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**:2262-2275.
- [7] CHENG Hong, WANG Zhi-Qiang, ZHANG Yao-Yu. Research on geometric rectification of aerial images [J]. *Journal of Northeast Normal University* (程红,王志强,张耀宇.东北师大学报), 2009, **41**(3):50-54.
- [8] Kennedy J, Eberhart R. Particle swarm optimization [J]. *Neural Networks, Proceedings of IEEE*. 1995, **4**:1942-1948.
- [9] Shi Y, Eberhart R. A modified particle swarm optimizer [J]. *Evolutionary Computation, Proceedings of IEEE*. 1998, **5**:69-73.

(上接 170 页)

[4] Serpico S, Bruzzone L. A new search algorithm for feature selection in hyperspectral remote sensing images [J] *IEEE Trans. Geosci. Remote Sens.*, 2001, **39**(7):1360-1367.

[5] Chang C-I, Wang S. Constrained band selection for hyperspectral imagery [J]. *IEEE Trans. Geosci. Remote Sens.*, 2006, **44**(6):1575-1585.

[6] Chang C-I, Du Q, Sun TL, *et al.* A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification [J] *IEEE Trans. Geosci. Remote Sens.*, 1999, **37**(6):2631-2641.

[7] Qian Y, Yao F, Jia S. Band selection for hyperspectral imagery using affinity propagation [J] *IET Comput. Vis.*, 2009, **3**(4):213-222.

[8] Frey B J, Dueck D. Clustering by passing messages between data points [J]. *Science*, 2007, **315**:972-976.

[9] Kullback S, Leibler R A. On information and sufficiency [J] *Ann. Math. Stat.*, 1951, **22**:79-86.

[10] Kullback S. *Information theory and statistics* [M]. New York: John Wiley and Sons, 1959.

[11] Cover T, Hart P. Nearest neighbor pattern classification [J]. *IEEE Trans. Inf. Theory*, 1967, **13**(1):21-27.

[12] Du Q, Chang C-I. A linear constrained distance-based discriminant analysis for hyperspectral image classification [J]. *Pattern Recognition*, 2001, **34**:361-373.

[13] Camps-Valls G, Bruzzone L. Kernel-based methods for hyperspectral image classification [J] *IEEE Trans. Geosci. Remote Sens.*, 2005, **43**(6):1351-1362.

[14] Melgani F, Bruzzone L. Classification of hyperspectral remote sensing images with support vector machines [J] *IEEE Trans. Geosci. Remote Sens.*, 2004, **42**(8):1778-1790.

[15] Landgrebe D. *Multispectral data analysis: A signal theory perspective* [R]. West Lafayette: School of Electrical & Computer Engineering, Purdue University, 1998:56-89.