

多尺度形态滤波及弹性匹配技术在类星体谱线识别中的应用

吴永东 马颂德

(中国科学院自动化所模式识别实验室 北京 2728 信箱 100080)

摘要 本文提出了一种类星体谱线证认方法。首先针对特征为极值点的信号,研究了多尺度膨胀(腐蚀)关于极值点数的两种重要特性及其应用。其一是单调率特性,根据它自动选择滤波器尺度,有效地滤除脉冲噪声;另一种是单调性,它是“从粗到精”策略来重新恢复极值特征位置的理论基础。根据这些性质,对光谱进行多尺度膨胀(腐蚀)和特征恢复,以滤除脉冲噪声而不影响谱线特征。然后研究弹性匹配技术应用于谱线证认,并指出了匹配方法中参量的物理意义。该方法对其他一些应用领域也行之有效。

关键词 类星体,单调性,单调率,弹性匹配

1 概述

在许多观测数据(如天体光谱、地震波)中,突出的峰值是被研究对象的重要特征,而脉冲噪声与信号特征相似,如何有效的抑制这类噪声是正确分析数据的前提条件。中值滤波器是抑制脉冲噪声的一种有效手段,它简单易行而且不影响峰值位置,可惜它的滤波效果与噪声强度有关,也就是说,难以选择合适的滤波器宽度。形态滤波器对输入对象的几何特征进行局部非线性变换,根据信号与噪声在不同尺度空间的特性差异,对脉冲型噪声有很好的滤波效果。由于不同脉冲型噪声的宽度不一样,大尺度滤波器常引起峰值移动,小尺度的滤波器难以滤除较大宽度的噪声,遗憾的是如何确定滤波器的尺度没有通用的方法。为了克服这一困难,我们采用了多尺度形态滤波。在滤波核为圆盘条件下,多尺度膨胀与腐蚀算子具有两个重要的特性:(1)单调性—随着尺度增加,极值点数单调减少;(2)单调率—极值点数减少速度几乎处处是减少的。根据特性2,存在某一尺度,所有大于该尺度的滤波器,对于滤波效果变

化不明显。由于信号与噪声能量(或宽度)存在较大的差别,在这一尺度下,信号能得到保存而大多脉冲型噪声被滤除。根据特性1,如果我们选择极值点为信号特征,在大尺度存在某一特征,则在小尺度依然存在该特征,当然特征位置可能有变化。按照“从粗到精”[1]策略进行跟踪,则原始特征位置可以恢复。一般说来,信号特征常常与研究的目的有关,要结合某一专业领域知识。类星体是宇宙中活动非常剧烈、具有极大红移值(这就是说,如果天体运行速度为 v ,真空中光速为 c ,则红移值 $z=(v-c)/c$)的天体,它是解决许多宇宙问题的关键。类星体的光谱是一些连续谱和发射谱(突出部分)组成(如图1a),在光谱曲线图上是发射谱(图1a中在70(90之间)远高于连续谱,并且有一定宽度,它是类星体的光谱特征。由于测量仪器与观测环境的限制,并非天体所有光谱都能观测到,只有一部分光谱能被地面测量仪器接受,并且天体在高速运动,根据多谱勒效应,观测的光谱存在较大的红移。所以弹性匹配方法适于类星体谱线证认:以典型类星体静止光谱为参考模板,对观测光谱进行弹性匹配[2,3],根据匹配结果确认发射线。

本文在第2节阐述了腐蚀算子的单调特性及其单调率,并且提供了其滤波尺度选择方法及特征恢复策略,第3节提供了弹性匹配选择类星体发射线的方法,第4节是实验及结果,最后是一简短结论。

2 多尺度形态滤波器

膨胀与腐蚀是最基本的滤波算子,其余算子(如形态开算子)是它们的复合。如果设输入信号 $f(x)$, 结构元为 $g_\sigma(x) = |\sigma|g(|\sigma|^{-1}x)$, 其中 σ 为尺度, 则它们的定义为

膨胀: $(f \oplus g_\sigma)(x) = \sup_t (f(x-t) + g_\sigma(t))$

腐蚀: $(f \ominus g_\sigma)(x) = \inf_t (f(x+t) - g_\sigma(t))$ 其中 t 属于两个函数定义域的交集

当结构为圆盘 $g_\sigma(x) = |\sigma|(\sqrt{1 - \|\sigma^{-1}x\|^2} - 1)$, $\|x\| \leq |\sigma|$ 时,膨胀(腐蚀)可视为一个圆盘从曲线上下方滚过时圆心的轨迹。显然圆半径越大,滤波后的曲线越平坦[4],但对原始特征的畸变也越明显。在实际应用中,如何确定圆半径没有一致的方法,一般根据先验知识或局域方差,显然,这两种方法存在很多局限。这里,根据多尺度膨胀(腐蚀)的单调性和单调率特性,采用一种较合理的尺度选择准则。我们考虑多尺度膨胀(腐蚀)单调率—极值点数减少速度几乎总是越来越慢,在尺度足够大时,极值点数减少速度接近于0。关于单调率特性可以直观解释如下:

设 $\sigma_1 > \sigma_2$, σ_1 和 σ_2 充分大,同时 $\Delta\sigma$ 充分小, $\sigma_1 = \sigma_2 + \Delta\sigma$, 则 $f \oplus g_{\sigma_1} \approx (f \oplus g_{\sigma_2}) \oplus g_{\Delta\sigma}$ 当经过大圆 σ_2 膨胀后,曲线已经足够光滑,小圆 $\Delta\sigma$ 膨胀已几乎不改变极值点数目。也就是说,增大滤波尺度并不总能改善对脉冲噪声的滤除效果。由于信号特征与脉冲噪声能量不一样,前者在各个尺度空间都保存,而脉冲型噪声只在一定的尺度空间存在。因此,根据膨胀(腐

蚀)关于极值点数的单调率特性,我们可以确定合适的滤波尺度 σ_0 , 当尺度大于 σ_0 时,极值点数减少不多。这样,在 σ_0 下选择的特征个数是最大可能的特征数。例如在图1中, (1. b) 是 (1. a) 在不同尺度下的极值点数关系图,从图中可以看出,极值点数具有单调性,而且当尺度大于40时,极值点数基本不变,因此尺度 σ_0 可选为40。

经过形态滤波后,特征(极值点)的位置可能发生移动,必须恢复特征位置。由于多尺度膨胀关于极值点数具有单调性,也就是说,在大尺度存在某一特征,则在小尺度依然存在该特征,因此在零尺度得到的位置是真正的特征位置。我们采用“从粗到精”的策略来跟踪其轨迹,逐步减少尺度,对于极值点的位置进行重新确定,直到尺度0。

3 弹性匹配证认谱线

3.1 匹配原理

设有两个集合:参考集 $\Pi = \{\lambda^i | i=1, 2, \dots, n\}$ 及测试集 $\Sigma = \{\hat{\lambda}^i | i=1, 2, \dots, m\}$, $\Sigma \subseteq (\lambda_1, \lambda_n)$, 任意选择 $\lambda \in \Sigma$, 置循环计数器 $i = 1$ 。

(1) 确定参考集空间

令 $\lambda \in \Sigma$ 与参考集 Π 中的某一个元素 λ^i 对应,记伸缩因子 $z = \lambda/\lambda^i$,再令集合 Φ 为空集。

$\forall \lambda_c \in \Pi$, 如果 $\lambda_1 \leq z\lambda_c \leq \lambda_n$ 则 $\Phi = \Phi \cup \{\lambda_c\}$

记集合 Φ 中元素个数为 n_i

(2) 赋匹配权值

依次搜索测试集中的每一个元素,测试它与参考集中的元素匹配程度,并根据测试结果赋予不同的权值。

$\forall \lambda_j \in \Phi, j = 1, 2, \dots, n_i$ 如果 $\exists \hat{\lambda} \in \Sigma, |z\lambda_j - \hat{\lambda}| \leq \epsilon_1, \epsilon_1$ 很小则说明匹配,赋予权值 $W_{ij} = 1.0$,

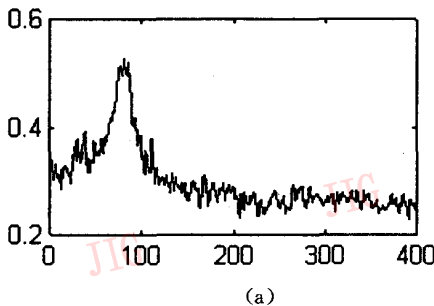
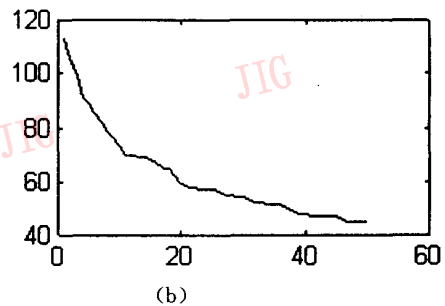


图1 (a) 原始曲线



(b) 极值点数与滤波尺度(的关系)

Fig. 1 (a) original curve (b) the relationship between the number of extreme and the scalar

否则 $W_{ij} = -0.5$ 总的权值为 $W_i = \sum_{j=1}^{n_i} W_{ij}$, 显然 $W_i \leq n_i$

(3) i 加 1, 重复 (1) 和 (2), 直到 i 等于 $n + 1$ 。

(4) 选择最佳匹配

如果 $W_k = \underset{i=1,2,\dots,n}{Max} W_i$, 则 λ 与 λ^k 之间的对应是最佳的, 然后依次确定其余匹配的谱线。

3.2 弹性匹配确定类星体发射线

在典型类星体光谱中, 有一些重要且易于检测的发射线。如氢巴尔末系 (H_α, H_β 和 H_γ) 及赖曼系 (L_{y_α} 和 L_{y_β}) 等, 它们构成集合 Π 。在所观测的天体光谱中, 选择的发射区构成集合 $\Sigma = \{\lambda | i = 1, 2, \dots, m\}$ 。对于一个类星体, 总有某一发射区具有高强度和大的宽度, 该区可选为匹配的起点 (即其波长选为算法第一步中的 λ), 从而加快匹配速度。另外, 天体运动导致观测到的光谱红移, 红移值正好对应弹性匹配算法中的伸缩因子。

在上述匹配算法中, 如果 W_k 太小, 就说明很少有被检测到的发射线, 此被测天体可能不是类星体。从理论上说, 一个红移值为 $z-1$ 的类星体, 集合 Φ 中所有元素应属于 Σ 。也就是说, $W_i = n_i$ 而不是 $W_i \leq n_i$ 。实践表明, 具有连续谱且表明被测天体是类星体。

4 实验结果

4.1 实验 1

图 2 是一个天体的光谱曲线, 波长单位是埃 (10^{-10} 米)。它的光谱是连续谱, 随着波长加大, 辐射强度总的趋势变小; 另外还有两个发射带。图 2 中直线是连续谱的逼近, 相邻两条垂直虚线标记发射带

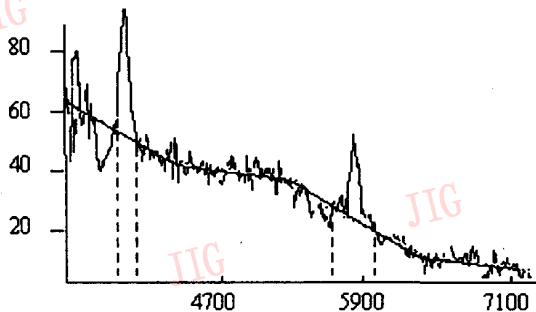


图 2 某天体的紫外光谱
Fig. 2 A stellar spectrum

的起止波长。经证认, 这两个发射线是 $C \text{ II}$ 和 $Mg \text{ I}$ 。它们的静止波长分别为 1909 埃和 2798 埃。表 1 列出了实验结果。第一列是光谱线名字; 第二列是发射带的起始波长; 第三列是发射带的终止波长; 第四列是发射带的表示波长, 用这一值来表示该发射区的波长; 第五列是天体红移值。

表 1

	起点波长	终点波长	代表波长	红移值
$C \text{ II}$	3940	4090	3999.5	1.0951
$Mg \text{ I}$	5716	6016	5847.5	1.0899

4.2 实验 2

图 3 是另一个天体的光谱曲线, 它有三个发射带, 图中直线段是连续谱的逼近值, 相邻两条垂直虚线标记发射带的起止波长。经证认, 这三个发射线是 $Mg \text{ I}$ 、 H_γ 和 H_β 。他们的静止波长分别为 2798 埃、4340 埃和 4861 埃。表 2 列出了实验结果, 表中各项含义同表 1。

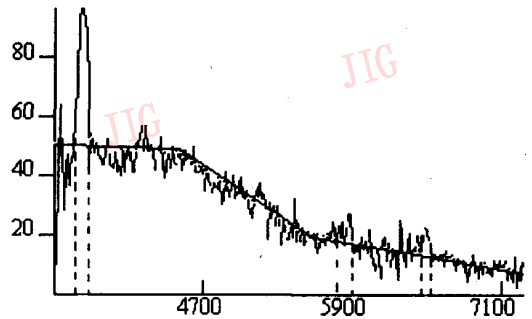


图 3 某天体的紫外光谱

Fig. 3 A stellar spectrum

表 2

	起点波长	终点波长	代表波长	红移值
$Mg \text{ I}$	3664	3796	3732.65	0.3340
H_γ	5710	5908	5827.68	0.3428
H_β	6418	6538	6480.32	0.3330

5 结论

多尺度分析是一种富有生命力的信号处理方法, 目前受到越来越多的重视。当结构元为圆盘时, 多尺度膨胀与腐蚀对于极值点数具有单调性, 这就适于对遭受强脉冲型噪声干扰的信号滤波, 也适于按照“从粗到精”的逆向过程确定特征的位置。在本文中, 我们不仅考虑多尺度膨胀的单调性, 而且研究其单调率, 并根据单调率几乎处处呈减少的趋势来

自动选择滤波器的尺度。同时,针对类星体的光谱特征,采纳弹性匹配技术证认谱线,对匹配算法中参数进行物理解释。事实上,上述方法适于许多其他领域(如心电图等)。

参考文献

- 1 A. P. Witkin, Scale-space Filtering, Proceeding of the 8th International Joint Conference on Artificial Intelligence, San Mateo, Calif, 1983,1019~1022
- 2 Wen-Hsiang Tsai and Shiao-Shian Yu, Attributed String Matching With Merging for Shape Recognition, IEEE Trans. On PAMI 1985,7(4):453~462
- 3 Xu-hong Xiao, Ru-wei Dai, A Hierarchical On-line Chinese Signature Verification System, 3rd ICDAR, Montreal, Canada, 1995, 202~205
- 4 Paul T. Jackay, et. al., "Scale-space Properties of the Multiscale Morphological Dilation-Erosion," IEEE Trans. on PAMI, 1996, 18(1): 38~51



吴永东 北京航空航天大学硕士毕业, 现于中国科学院自动化所攻读博士学位, 主要研究方向有模式识别、计算机视觉、过程控制及动态仿真。

Application of Multiscale Morphological Filtering and Elastic Matching on Quasar Spectrum Recognition

Wu Yong Dong, Ma Song De

(National Lab. of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences)

Abstract In this paper, a method about Quasar spectrum recognition is proposed. Firstly, on the signal with extreme points as its features, we introduce two important properties of multiscale morphological dilation(erosion). One is the monotone rate which is the basis of an adaptive scalar selection mechanism; the other is the monotone property which can be used to localize the features accurately by "Coarse to Fine" strategy. According to these properties, the spectrum is filtered by multiscale dilation and its features are recovered. Secondly, elastic matching is proposed to feature verification and its physical meaning is pointed out. In fact, this method can be applied to other fields effectively.

Keyword Quasar, monotone property, monotone rate, elastic matching