

# 基于 Mexican hat 函数的图像特征提取和配准算法

靳 峰 冯大政

(西安电子科技大学雷达信号处理国防重点实验室 西安 710071)

**摘 要** 使用 Mexican hat 函数作为特征检测算子,对图像进行局部区域提取和特征点提取,在此基础上,提出了一种结合局部区域和特征点的图像配准算法。使用 Mexican hat 检测算子和零交叉点检测的方法获得局部区域特征并进行初步匹配,然后使用基于不同尺度空间的 Mexican hat 检测算子进行特征点提取。将特征点按照局部区域进行分组,再对每一组内的特征点进行匹配操作。最后使用基于分组的随机采样一致性检验进行变换矩阵求解。算法使用 Mexican hat 特征检测算子进行两种图像特征的提取,对两种特征分别进行匹配,完成图像配准操作。实验结果表明,给出的 Mexican hat 特征检测算子在配准精度上不亚于当前主流检测方法,配准算法具有复现率高和运算速度快的优点。

**关键词** Mexican hat, 特征检测, 特征匹配, 图像配准

**中图分类号** TP391.4 **文献标识码** A

## Image Feature Detection and Registration Algorithm Based on Mexican hat Function

JIN Feng FENG Da-zheng

(National Key Laboratory of Science and Technology on Radar Signal Processing, Xidian University, Xi'an 710071, China)

**Abstract** An operator based on Mexican hat function was used for image local area and feature point detection. Then an image registration algorithm using the two kinds of features was proposed. The Mexican hat operator combining zero-crossing is used for local areas detection, and the feature points are detected by the operator on the different scale space. The image is partitioned into several regions by the local areas and the regions are matched. Then the points are grouped by the regions and matched in each group respectively. At last the image transaction function is gotten by the grouped random sample consensus. The algorithm in this work is based on two kinds of image feature detection and matching using the Mexican hat function, and the experimental results show that the proposed algorithm has high alignment accuracy and small computational volume.

**Keywords** Mexican hat, Feature detection, Feature matching, Image registration

## 1 引言

图像配准是指对取自不同时间、不同传感器或不同视角的同一景物的两幅图像或多幅图像进行匹配、叠加的过程,它是影像处理和分析的一个重要步骤,常常应用于多源遥感数据的融合分析、多时相遥感图像的变化检测和高程重建等方面。目前,采用基于图像特征提取的配准方法得到了广泛的应用,按照图像特征类型的不同,分为基于局部区域和基于特征点两种方法。

基于局部区域的典型算法有 Dai 等人提出的一种结合链码和不变矩的区域提取与配准算法<sup>[1]</sup>,该方法使用改进的 Laplace-of-Gaussian 算子进行图像的区域轮廓提取并计算其链码和不变矩从而进行图像配准。Tuytelaars 提出了一种局部仿射不变区域的概念<sup>[2]</sup>,该方法以图像滤波后的局部极大值为中心发散地向四周寻找局部极小值构成轮廓,然后以一个椭圆形区域代替该轮廓作为仿射不变区域。Matas 利用分

水岭算法提出一种最大稳定极值区域算法(MSER)<sup>[3]</sup>,通过用不同阈值分割得到的嵌套组找到最稳定的区域。国内的学者也做了大量的研究工作。陈秀新提出了一种基于连通区域的仿射不变区域提取算法<sup>[4]</sup>,唐涛等提出了一种基于关键点的闭合仿射不变区域提取算法<sup>[5]</sup>。

尺度空间理论提出后,在其基础上产生了很多图像特征检测和配准算法。Lowe 提出了一种称为 SIFT(Scale Invariant Feature Transform)的尺度不变配准算法<sup>[6]</sup>;Mikolajczyk 将尺度自适应 Harris 算子和自动尺度选择技术相结合,提出了尺度不变 Harris 算子<sup>[7]</sup>。围绕这两种方法产生了很多尺度不变和仿射不变的图像配准策略,国内的很多学者也进行了大量的研究工作。唐朝伟等提出了一种改进的 RIT-SIFT 算子<sup>[8]</sup>,范志强提出了一种基于聚类的 SIFT 特征匹配算法<sup>[9]</sup>,程邦胜和黄帅等分别对尺度不变 Harris 算子的检测参数进行了研究<sup>[10,11]</sup>,此外还有大量研究工作<sup>[12-17]</sup>。

本文提出了一种基于 Mexican hat 函数的特征检测算子,

到稿日期:2013-04-18 返修日期:2013-06-16 本文受国家自然科学基金:高维小样本自适应阵列信号处理研究及应用(61271293/F010305)资助。

靳 峰 男,博士生,主要研究方向为图像处理、计算机视觉,E-mail:xdjinfeng@163.com;冯大政 男,教授,主要研究方向为雷达信号处理、雷达成像与后处理技术。

将其应用于图像的局部区域提取和特征点提取。在获得了两种图像特征之后,将特征点按照局部区域进行分组,分别进行局部区域匹配和每一个区域内的特征点匹配,最后使用随机采样一致性检验求出图像变换矩阵,完成配准操作。

## 2 基于 Mexican hat 函数的特征提取

### 2.1 Mexican hat 函数

Manjunath 指出,所有基于空间尺度理论的图像特征检测方法,其根本都是基于对不同带宽的低通滤波器的滤波响应的观察与对比<sup>[18]</sup>,基于尺度空间理论的图像信息描述和特征定位的归一化描述方法如式(1)所示:

$$Q(x, y) = f(W(x, y, s_1) - \gamma W(x, y, s_2)) \quad (1)$$

式中, $Q$ 为尺度空间检测算子, $W$ 为检测图像在离散尺度空间  $s$  上的滤波响应, $f$ 为调制函数, $\gamma$ 为常数。研究表明,高斯函数是唯一的空间核函数,当  $W$  取高斯核函数, $f$  取连续尺度空间差分,  $Q$  即为经典的高斯差分算子(DoG)。相邻尺度空间的高斯差分为:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) \otimes I(x, y) \quad (2)$$

式中, $\sigma$ 为尺度因子, $k$ 为构造尺度空间的尺度变化因子。由于高斯差分算子与高斯-拉普拉斯算子  $\nabla^2 G$  存在如下关系:

$$\frac{(G(x, y, k\sigma) - G(x, y, \sigma))}{k\sigma - \sigma} \approx \partial G / \partial \sigma = \sigma \nabla^2 G \quad (3)$$

式中,

$$\nabla^2 G(x, y) = -1/2\pi\sigma^4 (1 - (x^2 + y^2)/2\sigma^2) \exp(-(x^2 + y^2)/2\sigma^2) \quad (4)$$

因此,

$$G(x, y, k\sigma) - G(x, y, \sigma) \approx (k-1)\sigma^2 \nabla^2 G \quad (5)$$

由此得到一个新的尺度空间滤波函数  $R$ ,它是一个 Mexican hat 函数,本文称之为 Mex 算子。

$$R = -1/2\pi\sigma^2 (1 - (x^2 + y^2)/2\sigma^2) \exp(-(x^2 + y^2)/2\sigma^2) \quad (6)$$

Mex 算子与  $\nabla^2 G$  算子和 DoG 算子的二维波形如图 1 所示。可以看到三者非常接近,只在平滑度上有所差异。DoG 算子是目前流行的特征提取策略 SIFT 算法的基础<sup>[4]</sup>,  $\nabla^2 G$  算子在图像特征提取领域有着广泛的应用。由于 Mex 算子和 DoG、 $\nabla^2 G$  存在上述推导关系,同时 Mexican hat 小波函数在图像特征提取上的效果也得到了验证<sup>[19]</sup>,本文采用 Mex 算子作为特征检测算子进行图像区域特征和点特征的提取。这样做的优势在于, Mexican hat 函数可以认为是高斯函数的二阶导<sup>[20]</sup>,它的局部极值点也就是高斯差分尺度空间上的极值点。因此,使用 Mexican hat 函数代替高斯函数进行尺度空间上的特征检测,可以简化运算。

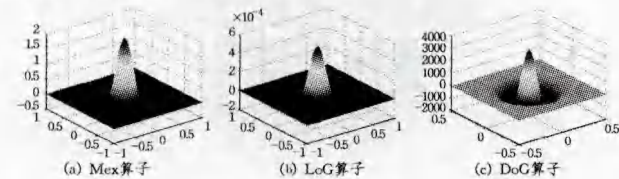


图1 二维波形

### 2.2 局部区域提取

零交汇点检测是图像边缘检测上较为常用的方法。该方法利用图像灰度变化在一阶导数和二阶导数上的特性来实现图像特征提取,二阶导数与数轴交汇的点反映了图像在灰度上的剧烈变化,因此可以在此基础上探测具有显著特征的像素点。

$$f(x, y) = D^2 G \otimes I(x, y) \quad (7)$$

$D^2$  为二阶求导运算,  $G \otimes I(x, y)$  为使用 Gaussian 函数滤波对图像进行平滑处理。由于图像中某一方向灰度变化的二阶导数只有在其垂直方向上才能取得最大曲率,因此零交汇点检测算法存在二阶导数的方向选择问题。为了解决这一问题,可使用旋转不变的检测算子来代替  $D^2$ ,如著名的 LoG 检测算法使用  $\nabla^2 G$  算子<sup>[21]</sup>。本文使用 Mex 算子来实现对零交汇点检测的改进。由图 1 可知, Mex 算子是不依赖于方向的;而且, Mexican hat 函数其分布无论在时域上还是在频域上都是旋转不变的<sup>[21]</sup>。使用 Mex 算子对零交汇点检测进行改进,得到的特征检测算法和 LoG 算法可以达到几乎一致的效果,如图 2 所示,尺度因子  $\sigma=0.02$ 。



图2 边缘提取效果对比

### 2.3 特征点提取

将 Mex 算子代入式(1),可得到基于 Mexican hat 函数的尺度空间上的特征点检测方法。

$$Q(x, y) = |(R(x, y, s_1) - R(x, y, s_2)) \otimes I(x, y)| \quad (8)$$

式中,  $\otimes$  表示卷积运算,  $s$  为尺度变化因子,  $s = -\log_2 \sigma$ 。  $Q(x, y)$  的极值点即为所求的特征点。调制函数  $f$  选取的是取绝对值运算,因为在特征检测中,我们要得到的是局部最大值或者最小值,而不关心极值的正负。

图 3 描述了不同尺度空间的滤波效果和局部极大值检测,从左到右分别为尺度变化因子  $s$  取值为 1, 2, 3 和 4 时的检测结果。



图3 尺度空间检测

## 3 图像配准

### 3.1 分组匹配策略

基于 Mex 算子在图像特征检测上的效果,本文提出了一种结合区域特征和点特征的图像配准算法。将特征按照某种形式进行分组是一种能够提高特征匹配效率的方法,本文使用局部区域对图像进行划分,在此基础上进行特征点的分组和匹配。首先进行局部区域的匹配,局部区域特征数量少,运算速度快;然后,为了进一步提高配准精度,按照局部区域对图像进行划分,将特征点按照图像划分的范围进行分组,进行组内的特征点匹配。每组内的特征点数量相对较少,运算速度快,同时,避免了不同组直接的特征点发生误匹配。具体步骤如下:

1)对图像的局部区域特征进行提取,包含 Mex 算子滤

波、极化处理和零交汇点检测。

2)使用式(8)进行尺度空间上的特征点提取。

3)以区域重心为圆心、 $r_i$  为半径将图像分为  $n$  个圆形区域,落在圆内的特征点随之分为  $n$  组;舍弃落在圆外的特征点。 $r_i$  为区域  $i$  闭合线上所有点到重心平均距离的 1.2 倍。

4)进行局部区域匹配和区域内的特征点匹配,如图 4 所示。

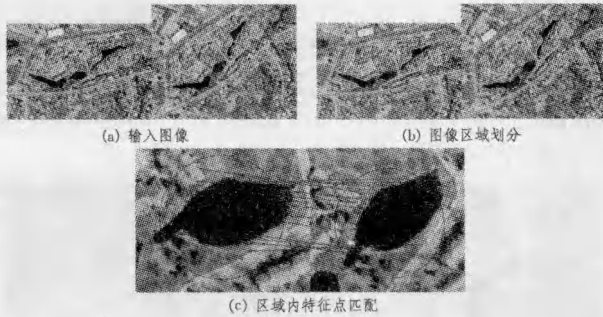


图 4 特征匹配

图 4(b)给出了一组圆形局域内部的特征点匹配。使用归一化的圆形区域是对图像区域划分的简化处理,这样做有两个优点:1)降低了特征点分组判断的复杂性,因为特征点是否位于区域内部只需比较其到圆心的距离与圆的半径即可;2)降低了局部区域抽取精度对分组配准的影响,因为没有使用区域边缘而使用边缘到区域重心的均值作划分的边界。除非是局部区域匹配发生匹配错误的情况,否则区域提取精度对特征检测和匹配的影响被大大降低。而对于局部区域匹配发生错误的情况,下文使用随机采样一致性检验的方法来解决。

由于特征检测和匹配的时间复杂度为  $O(l_{sum}D)$ ,其中  $l_{sum}$  为特征数量, $D$  为特征描述向量的维度,因此,当特征描述子不变时,在满足一定有效特征数量的前提下,特征数量越少,匹配效率越高。在特征点分组匹配的过程中,匹配运算的时间复杂度为  $O(l_{max}D)$ ,其中  $l_{max}$  为最大分组的特征点数量,显然  $l_{max} \leq l_{sum}$ ;同时,所有检测到的特征点都参与匹配运算,有效特征点数量没有减少。

### 3.2 特征匹配算法

由于局部区域特征和区域内的特征点数目比较少,本文使用简单的投票策略进行特征匹配。投票策略是一种基于距离的内点选取策略<sup>[22]</sup>,其机制如图 5 所示。

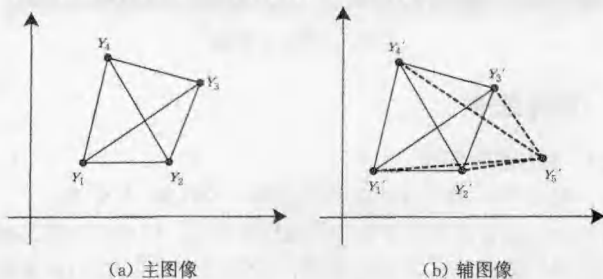


图 5 样本空间均值点定位

主图像中的点  $Y_4$  到周围点的距离集合为  $\{dis_{4,1}, dis_{4,2}, dis_{4,3}\}$ ,辅图像中的点  $Y_4'$  到周围点的距离集合为  $\{dis'_{4,1}, dis'_{4,2}, dis'_{4,3}, dis'_{4,5}\}$ ,这两个集合中有 3 个元素值是相等的,占前者的全部和后者的 3/4,认为这两个点是一对匹配点。辅图像中的点  $Y_5'$  到周围点的距离集合,在主图像中找不到与之近似的集合,认为该点是一个外点。投票策略的公式化描述为:

$$V(u, v) = \sum_{p \in [1, M_1]} \sum_{q \in [1, M_2]} \exp\left(-\left(1 - \frac{H(u, p)}{H'(v, q)}\right)^2\right) \quad (9)$$

其中,  $M_1$  和  $M_2$  分别为主轴图像待匹配点的个数,  $H$  和  $H'$  分别为主轴图像中所有待匹配点间的距离构成的矩阵,  $V$  为投票矩阵。若  $V(u, v)$  是矩阵  $V$  在第  $u$  行和第  $v$  列的最大值,并且  $V(u, v) \geq T$ ,  $T$  为指定常数,那么可以认为主图像中第  $u$  个点与辅图像中第  $v$  个点是一对匹配的特征点。

### 3.3 求解变换矩阵

最后,本文采用在分组基础上的随机采样一致性检验进行外点删除和图像变换矩阵求解。具体步骤如下:

1)分别计算  $n$  组由匹配的特征点对得到的变换矩阵  $\Gamma$  和均方根误差  $R$ 。

2)随机选取一个变换矩阵  $\Gamma$  检验所有特征点,求出满足  $\|\Gamma \cdot \bar{X} - \bar{X}'\| < \epsilon$  的点集合  $\Pi$ ,  $\Pi$  的大小为  $sum$ ,  $\bar{X}$  和  $\bar{X}'$  为主副图像中特征点坐标,  $\epsilon$  为容错误差。若  $sum$  大于指定常数,则令  $\Pi$  作为内点集合,返回 1); 否则,返回 2)。

3)经过  $\lambda$  次迭代,求出满足一定容错误差的最大特征点对集合,得到稳定的最小均方根误差的变换矩阵。

上述过程的时间复杂度为  $O(C_T)$ ,  $C_T$  取决于步骤 2) 随机选取符合条件的变换矩阵  $\Gamma$  的概率。该求解策略采用随机采样一致性检验,不仅可以有效删除外点,同时也解决了局部区域匹配发生错误时的特征点匹配问题。因为只要有一个组的特征点能够完成匹配,就能够在整个图像中获得最大数量的匹配点对。在区域匹配完全失败,即一个局部区域也没有匹配成功的情况下,迭代次数达到上限  $\lambda$  仍不能求出满足容错误差的特征点对集合,算法崩溃。

## 4 实验分析

图 6 为采用不同拍摄角度下的自然光学图像的匹配效果。图 6(a)为不同角度拍摄的原始光学图像,图 6(b)和图 6(c)分别为局部区域和特征点的提取,由于图像较大,仅将部分区域的特征匹配情况单独列出,如图 6(d)所示。

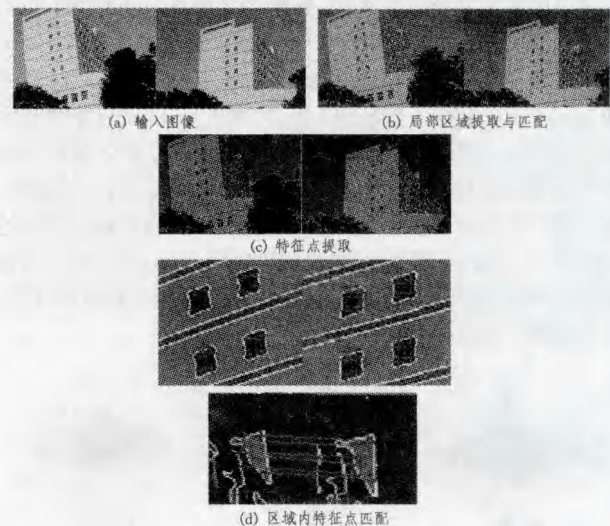


图 6 图像匹配

经过局部区域划分和匹配操作后,主轴图像分别被划分为 20 个区域进行特征点匹配,如图 6(b)和图 6(d)所示。将本文算法与两种典型的图像配准算法——基于局部区域检测和匹配的 LoG 算法和基于特征点的 SIFT 算法进行比较,在复现率、配准精度和配准效率 3 个方面的比较结果如表 1 所列。本文实验的硬件环境为 CPU2.1GHz,内存 2GB,操作系

统为 win7, 程序开发平台为 matlab2012。作为对比的 LOG 算法采用 matlab2012 的内置函数, SIFT 算法程序来自于文献[6]和作者发布的 matlab 程序数据。投票策略参数  $T=0$ , 随机采样一致性检验参数  $\epsilon=0.8$ 。

表 1 配准结果

	特征数目	复现率	RMSE	运行时间
本文方法	1078	72%	0.0726	120s
SIFT	1204	48%	0.0634	513s
LoG	73	/	0.7430	13s

表 1 中的复现率表示匹配成功的特征数量占全体特征数量的百分比, 复现率越高, 说明特征检测的性能越稳定, 图像配准的准确率越高。配准精度用均方根误差 (Root Mean Square Error, RMSE) 衡量, 配准效率用运行时间衡量。LoG 算法使用的是能够提供最稳定图像特征的检测算子<sup>[18]</sup>, 它获取的图像特征数目少且定位较准确, 比较复现率意义不大。本文使用的检测算子在图像边缘检测上的效果与其相当, 但是本文算法在图像配准上取得的精度超过 LoG 算法, 这是因为本文算法使用的是基于特征点的配准方法, 该方法在有效特征数量上高于 LoG 算法。另外, 本文算法在配准精度上达到了与目前比较流行的 SIFT 算法相当的程度, 但是由于采取了分组策略, 运算时间大大缩短。经过 10 次重复检验, 获得的结果与表 1 接近, 表明本文算法在图像特征检测和匹配上兼有两种典型算法的优点。

为了进一步验证算法, 使用 3 组公共测试图像和 2 组实际拍摄图像进行配准算法的验证, 如图 7 所示。图像涵盖尺度变换、大角度旋转、光照变化和图像的仿射变换。



图 7 检测图像

将本文算法与改进的 SIFT 算法<sup>[8]</sup>和尺度 Harris 算法<sup>[11]</sup>进行对比, 在复现率和运行时间上的效果如图 8 所示。图 8(a)和图 8(b)中横坐标 1 到 5 分别表示图 7 检测图像中的图 7(a)到图 7(e)。在光照条件、旋转和尺度变化下, 本文算法都取得了较好的效果。由于本文算法使用的仍然是基于尺度空间理论的高斯函数滤波, 它对于光照条件和尺度变化是不敏感的, 如图 7(d)和图 7(e)所示。另外, 本文算法在匹配操作中使用的形状、轮廓和距离等几何特征, 在旋转变化下也能保持较好的不变特性, 如图 7(a)和图 7(b)所示。因此, 在基本的图像变化中, 3 种算法都能保证一定的复现率, 但是本文算法复现率最高, 而且在运算时间上具有明显优势, 如图 8(a)和 8(b)所示。原因在于本文算法采用了不亚于 SIFT 算子的特征检测方法, 但是在特征描述和特征匹配上采用了更加有效的方法。在图像发生仿射变换的情况下, 如图 7(c)所示, 3 种算法的复现率都有所下降; 本文算法在特征匹配过程中使用仿射不变的 Mahalanobis 距离代替常用的欧式距离, 增加了运算时间。

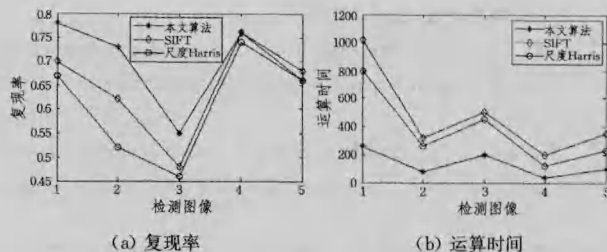


图 8 对比效果

文献[9]没有对 SIFT 算子进行改进, 而是在特征匹配上提出了新的方法; 文献[11]对尺度 Harris 算子的参数进行了详细分析, 给出了能够得到更多关键点的参数空间。这两种算法代表了基于特征的图像配准算法的典型研究方向。SIFT 算法在特征检测和特征描述上具有较高的精度, 但是运算时间长, 复现率较低。文献[9]使用聚类和分段匹配的方法, 一定程度上提高了复现率, 但是在运算时间上仍然不具有优势, 如图 8(a)和 8(b)所示。文献[11]给出了一个合适的参数区间, 使得尺度 Harris 算子能够像 SIFT 算子一样提供足够多的关键点, 但是缺少有效的特征匹配方法, 复现率较低, 如图 8(a)所示。

相对而言, 本文算法在复现率和运算时间上具有优势, 尤其是在图像像素较多、图像内纹理特征较为复杂的情况下, 如图 7(a)和图 7(c)所示。这是因为, 本文采用的分组策略在时间复杂度上具有明显优势; 而且, 基于局部区域和基于特征点的分段匹配策略保证了复现率, 这与文献[9]的分段匹配思想类似。最后, 用于特征描述和特征匹配的方法是基于几何特征的, 基于几何特征或几何关系的匹配方法在特征数量较大的情况下具有复现率和运算时间上的优势, 目前已经成为图像处理领域的研究热点。

**结束语** 提出了一种基于 Mexican hat 函数的图像特征提取和配准算法。通过对尺度空间上的高斯差分算子的研究, 得到了一个 Mexican hat 函数的图像特征检测算子。使用该算子对图像进行局部区域和特征点的检测, 并使用分组策略, 将两种图像特征综合利用进行图像配准工作。最后使用随机采样一致性检验进行外点删除和图像变换矩阵求解。经过实际的图像配准验证, 本文给出的图像特征检测算子是有效的。本文给出的配准算法是一个快速、准确的配准算法。两种图像特征相结合的分组配准策略减少了配准运算的时间, 基于尺度空间上的特征点检测与匹配保证了配准的精度。

## 参考文献

- [1] Dai Xiao-long, Khorram S. A feature-based image registration algorithm using improved chain-code representation combined with invariant moment[J]. IEEE Transaction on Geoscience and Remote Sensing, 1999, 37(5): 2351-2362
- [2] Tuytelaars T, Gool L V. Wide baseline stereo matching based on local, affinity invariant regions[C]// Proceedings of the 11th British Machine Vision Conference. Bristol, UK: ILES Central, 2000: 412-425
- [3] Matas J, Chum O, et al. Robust wide-baseline stereo from maximally stable extremal regions[C]// Proceedings of the British Machine Vision Conference. Cardiff, UK, 2002: 384-393
- [4] 陈秀新, 贾克斌. 基于连通区域的仿射不变区域提取方法[J]. 计算机工程, 2011, 37(20): 18-20
- [5] 唐涛, 粟毅, 陈涛, 等. 一种新的图像局部仿射不变特征提取方法

- [6] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2), 91-110
- [7] Mikolajczyk K, Schmid C. Scale & Affine Invariant Interest Point Detectors[J]. *International Journal of Computer Vision*, 2004, 60(1): 63-68
- [8] 唐朝伟, 肖健, 邵艳清, 等. 一种改进的 SIFT 描述子及其性能分析[J]. *武汉大学学报: 信息科学版*, 2012, 37(1): 11-16
- [9] 范志强, 赵沁平. 一种基于数据聚类的鲁棒 SIFT 特征匹配方法[J]. *计算机研究与发展*, 2012, 49(5): 1123-1129
- [10] 程邦胜, 唐孝威. Harris 尺度不变性关键点检测子的研究[J]. *浙江大学学报: 工学版*, 2009, 43(5): 855-859
- [11] 黄帅, 吴克伟, 苏菱. 基于 Harris 尺度不变特征的图像匹配方法[J]. *合肥工业大学学报: 自然科学版*, 2011, 34(3): 379-382
- [12] 钟金琴, 檀结庆, 等. 基于二阶矩的 SIFT 特征匹配算法[J]. *计算机应用*, 2011, 31(1): 29-32
- [13] 张海燕, 李元媛, 储晨响. 基于图像分块的多尺度 Harris 角点检测方法[J]. *计算机应用*, 2011, 31(2): 356-357
- [14] 王鹏, 王平, 沈振康, 等. 一种基于 SIFT 的仿射不变特征提取方法[J]. *信号处理*, 2011, 27(1): 88-93
- [15] 于丽莉, 戴青. 一种改进的 SIFT 特征匹配算法[J]. *计算机工程*, 2011, 37(2): 210-212
- [16] 程德志, 李言俊, 余瑞星. 基于改进 SIFT 算法的图像匹配方法[J]. *计算机仿真*, 2011, 28(7): 285-289
- [17] 张海燕, 李元媛, 储晨响. 基于图像分块的多尺度 Harris 角点检测方法[J]. *计算机应用*, 2011, 31(2): 356-357
- [18] Manjunath B S, Shekhar C, Chellappa R. A new approach to image feature detection with application[J]. *Pattern Recognition*, 1996, 29(4): 627-640
- [19] Yasein M, Agathoklis P. A feature-based image registration technique for images of different scale[C]// *IEEE International Symposium on Circuits and Systems*. May 2008: 3558-3561
- [20] Daubechies I. *Ten Lectures on Wavelets*[M]. Fourth Printing, Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1992
- [21] Mikolajczyk K. Detection of local features invariant to affine transformations[D]. Institut National Polytechnique de Grenoble, France, 2002
- [22] Qian Wei, Fu Zhi-zhong, et al. Voting-strategy-based approach to image registration[J]. *Opto-Electronic Engineering*, 2008, 35(10): 86-91

(上接第 263 页)

及特定测序目标区域不同, 本文算法不仅能扩增特定的基因目标区域(如 16S/18S rRNA), 而且能有效模拟 ePCR 误差。

**结束语** 本文在现有的一些研究基础之上, 提出了一种新的 454 模拟算法, 它基于易错 PCR 模型模拟乳液 PCR 过程, 分别采用正态分布模型及对数正态分布模拟边合成边测序过程中的正信号流与负信号流过程。实验结果表明, 该算法具有较好的性能, 能在有限的时间内模拟大量的测序列, 所需要的时间与模拟序列数量成正比。与此同时, 该算法具有易推广至其它二代测序平台的优点, 通过引物局部序列比对获得待扩增区域, ePCR 扩增过程结合现有的一些算法, 如 ART 即可完成对基因特定目标区域的跨平台模拟测序。由上述分析可见, 该算法不仅能完成特定目标基因区域模拟测序, 帮助我们定量、定性分析微生物种群, 还有助于验证评价当前 454 测序数据分析软件, 促进二代测序数据分析。

### 参 考 文 献

- [1] Grice E A, Kong H H, Conlan S, et al. Topographical and temporal diversity of the human skin microbiome [J]. *Science*, 2009, 324(5931): 1190-1192
- [2] 蒋德明, 孙玉华, 李丹, 等. 基于 16S rRNA 基因序列分析受砷和硫酸盐污染的土壤细菌多样性 [J]. *微生物学通报*, 2011, 38(10): 1592-1601
- [3] Oakley B B, Fiedler T L, Marrazzo J M, et al. Diversity of human vaginal bacterial communities and associations with clinically defined bacterial vaginosis [J]. *Appl Environ Microbiol*, 2008, 74(15): 4898-4909
- [4] Kellenberger E. Exploring the unknown: The silent revolution of microbiology [J]. *Embo Rep*, 2001, 2(1): 5-7
- [5] Margulies M, Egholm M, William E, et al. Genome sequencing in microfabricated high-density picolitre reactors [J]. *Nature*, 2005, 437(7057): 376-380
- [6] 刘玮琦, 菲振川, 杨宇红, 等. 应用 16S rRNA 基因文库技术分析土壤细菌群落的多样性[J]. *微生物学报*, 2008, 35(10): 1344-1350
- [7] Ley R E, Backhed F, Turnbaugh P, et al. Obesity alters gut microbial ecology [J]. *PNAS*, 2005, 102: 11070-11075
- [8] 曹波, 杨红, 许强华, 等. 基于 16S rRNA 技术的长江口微生物分子生物学鉴定与分析 [J]. *上海大学学报*, 2011(2): 191-197
- [9] Benoit M R, Li W, Stodieck LS, et al. Microbial antibiotic production aboard the International Space Station [J]. *Appl Microbiol Biotechnology*, 2006, 70: 403-411
- [10] Richter D C, Ott F, Auch A F, et al. MetaSim — A Sequencing Simulator for Genomics and Metagenomics [J]. *PLoS ONE*, 2008, 3(10): e3373
- [11] Balzer S, Malde K, Lanzén A, et al. Characteristics of 454 pyrosequencing data — enabling realistic simulation with flowsim [J]. *Bioinformatics*, 2010, 26(18): i420-i425
- [12] Lysholm F, Andersson B, Persson B. An efficient simulator of 454 data using configurable statistical models [J]. *BMC Research Notes*, 2011, 4: 449
- [13] Huang Wei-chun, Li Le-ping, Myers J R, et al. ART: a next-generation sequencing read simulator [J]. *Bioinformatics*, 2012, 28(4): 593-594
- [14] 宣黎明, 韦朝春, 李亦学. 基于 GPU 运算的宏基因组二代测序模拟软件 [J]. *华东理工大学学报*, 2012(4): 472-476
- [15] Huse S M, Welch D M, Morrison H G, et al. Ironing out the wrinkles in the rare biosphere through improved OTU clustering [J]. *Environ Microbiol*, 2010, 12(7): 1889-1898
- [16] Sharpston mail T J, Samantha, et al. PhylOTU: A High-Throughput Procedure Quantifies Microbial Community Diversity and Resolves Novel Taxa from Metagenomic Data [J]. *PLoS Comput Biol*, 2011, 7(1): 1-13
- [17] Pritchard L, Corne D, Kell D, et al. A general model of error-prone PCR [J]. *J Theor Biol*, 2005, 234(4): 497-509
- [18] Wang Q, Garrity G M, Tiedje J M, et al. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy [J]. *Appl Environ Microbiology*, 2007, 73(16): 5261-5267