



南昌航空大学

毕业设计（论文）

题 目： 2-DE 图像匹配算法研究

学 院： 信息工程学院

专业名称： 通信工程

班级学号： 10042115

学生姓名： 王超

指导教师： 欧巧凤

二〇一四年六月

2-DE 图像匹配算法研究

学生姓名：王超

班级：100421

指导老师：欧巧凤

摘要：随着后基因组时代的到来，研究兴趣已经转移到蛋白质水平的监测。蛋白质组学已然成为后基因时代的主要研究内容，而 2-DE 图像间差异蛋白质点获取和分析是蛋白质组学研究的核心技术手段，而同源蛋白质点匹配是差异蛋白质点获取的关键环节。

本文以 2-DE 图像蛋白质点为研究对象，探索凝胶图像间同源蛋白质点的匹配方法。主要研究内容如下：首先，阐述了蛋白质组学的发展现状及二维凝胶图像的匹配方法。其次，通过手动添加特征点即 Landmark 标记点，提出了基于 Landmark 的蛋白质点匹配算法，该算法人工介入较多。接着，提出一种基于 SURF 特征点和 RANSAC 算法的蛋白质点自动匹配算法；在该算法中，首先通过 SURF 算法得到一系列包含异常点的特征点对，然后利用 RANSAC 算法从包含异常点的特征点对中，通过迭代方式估计匹配模型并去除异常点，从而实现匹配。最后，提出了一种基于几何相似性的匹配算法思路。

关键词：2-DE 图像，蛋白质点匹配，Landmark，SURF，RANSAC

指导老师签名：

Research on the Matching Algorithm of 2-DE Image

Student Name: Wang Chao

Class: 100421

Supervisor: Ou Qiaofeng

Abstract: With the advent of the post-genomic era, research interest has moved to monitoring on protein levels. Proteomics has become the main contents of the post-genome era. The detection and analysis of differential protein spots in the 2-DE images are the core techniques of proteomics research. The aim of 2-dimensional electrophoresis (2-DE) image match is to find homologous proteins and differential proteins.

This paper takes 2-DE image protein as the research object, exploring the matching method of homologous protein spots in the gel image. The main contents are as follows: Firstly, the current development of proteomics and the matching method of two-dimensional gel images was described in this paper. Secondly, the matching algorithm between proteins, which is based on Landmark, is proposed. The algorithm is proposed by adding the feature points manually, which works under more manual intervention. Then an automatic matching algorithm which is based on SURF feature points and RANSAC algorithm in protein points is proposed. Firstly, a series of feature points which contain outliers are obtained by the SURF algorithm. Then in order to achieve matching, a RANSAC algorithm which is in an iterative manner is used to estimate matching model and remove abnormal points in the feature point pairs which contain outliers. Finally, a matching idea which is based on geometric similarity is put forward.

Keywords: two-dimensional electrophoresis image, spot matching, landmark, SURF, RANSAC

Signature of Supervisor:

目 录

第一章 绪论.....	1
1.1 课题背景及研究意义.....	1
1.1.1 蛋白质组学简介.....	1
1.1.2 二维凝胶电泳技术简介.....	1
1.1.3 课题研究意义.....	2
1.2 图像匹配研究现状.....	2
1.3 本课题的主要工作流程及本文的主要内容.....	3
第二章 基于 Landmark 的蛋白质点匹配算法.....	5
2.1 概述.....	5
2.2 Landmark 标记点的选取.....	6
2.3 选取两凝胶图像间的几何映射模型.....	8
2.4 两 2-DE 图像间的蛋白质点匹配.....	9
2.5 实验结果及分析.....	11
2.6 本章小结.....	17
第三章 基于 SURF 和 RANSAC 的蛋白质点匹配算法.....	18
3.1 概述.....	18
3.1.1 SURF 算法.....	18
3.1.2 RANSAC 算法.....	21
3.2 2-DE 图像间蛋白质点匹配.....	23
3.3 实验结果及分析.....	24
3.4 本章小结.....	26
第四章 不同匹配算法的对比.....	27
4.1 基于几何相似性的蛋白质点匹配算法.....	27
4.2 三种不同算法的对比.....	29
第五章 总结.....	31
参考文献.....	33
致谢.....	34
附录.....	35

第一章 绪论

1.1 课题背景及研究意义

1.1.1 蛋白质组学简介

蛋白质组学 (proteomics) 一词最早由 Marc Wilkins 提出, 名称来自蛋白质 (protein) 与基因组学 (genomics) 两词组合^[1]。它以细胞内全部蛋白质的存在及活动方式为研究对象。双向凝胶电泳、等电聚焦、生物质谱分析及非凝胶技术等都是蛋白质组学研究中常用的技术手段。蛋白质组学本质上指的是在大规模水平上研究蛋白质的特征, 包括蛋白质的表达水平、翻译后的修饰以及蛋白与蛋白间相互作用等, 由此全面认知蛋白质水平上的疾病发生和细胞代谢等过程。

1.1.2 二维凝胶电泳技术简介

二维凝胶电泳 (又称双向凝胶电泳) 的原理是第一向基于蛋白质的等电点不同用等电聚焦分离, 第二向则按分子量的不同用聚丙烯酰胺电泳 (SDS-PAGE) 分离, 把复杂蛋白混合物中的蛋白质在二维平面上分开^[2]。如图 1.1 所示, 横向采用电流作用下形成 pH 梯度, 根据等电点的量进行分离; 纵向依据蛋白质的大小进行分离。二维凝胶电泳技术可用于蛋白质转录及转录后修饰研究, 蛋白质组的比较和蛋白质间的相互作用, 细胞分化凋亡研究, 致病机制及耐药机制的研究, 疗效监测, 新药物开发, 癌症研究, 蛋白纯度检查, 小量蛋白纯化, 新替代疫苗的研制等许多方面^[3]。二维凝胶电泳、质谱分析和生物信息学是蛋白质组学研究最基本、最主要的技术。二维凝胶电泳是目前唯一可以将数千种蛋白质同时分离的方法, 它与质谱技术联合起来使用被认为是蛋白质组研究技术的标准方法, 广泛应用于生物学研究的各领域。

从很多方面来说, 二维凝胶电泳是最直接寻找和预分离未知蛋白及其翻译后修饰形式 (post-translationally modified forms, PTMs) 的方法。假如一张二维凝胶电泳图像上的多个位点确认是同一种蛋白, 那么这种位点的迁移只能是用 PTMs 解释。二维凝胶电泳中常可发现蛋白拖曳现象, 这很可能是一个蛋白的不同翻译后修饰产物所造成的^[4]。例如, 二维凝胶电泳能够有效的观察肌动蛋白 M 带的蛋白磷酸化情况^[5]。

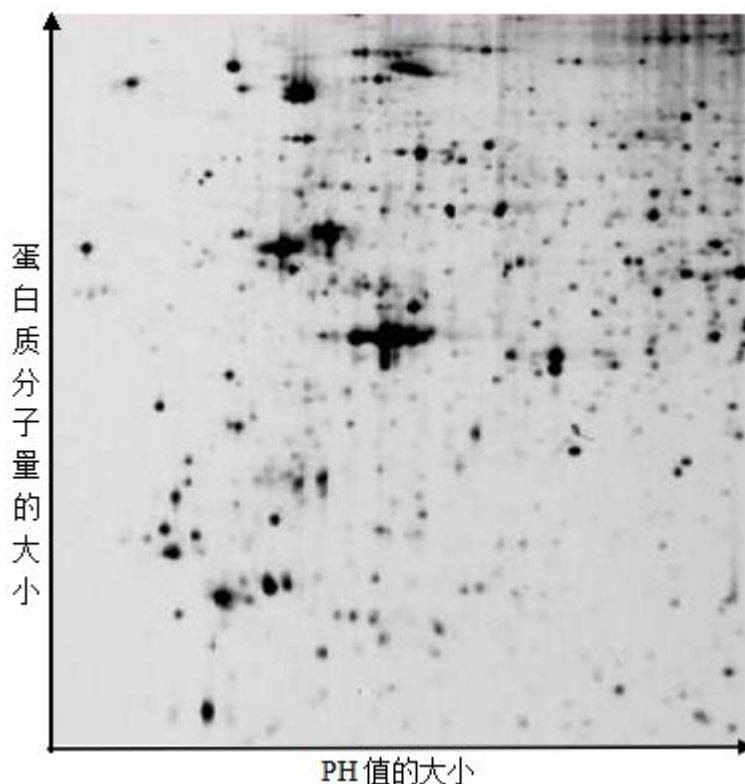


图 1.1 二维凝胶电泳图

1.1.3 课题研究意义

蛋白质组学的研究是生命科学进入后基因时代的特征，而 2-DE 图像分析技术是蛋白质组学研究的重要技术，是同源蛋白质点匹配差异蛋白质点提取的关键技术。蛋白质组学的研究不仅能为生命活动规律提供物质基础，也能为众多种疾病机理的阐明及攻克提供理论依据和解决途径。通过对正常个体及病理个体间的蛋白质组比较分析，可以找到某些“疾病特异性的蛋白质分子”，它们可成为新药物设计的分子靶点，或者也会为疾病的早期诊断提供分子标志。因此，蛋白质组学研究不仅是探索生命奥秘的必须工作，也能为人类健康事业带来巨大的利益。

1.2 图像匹配研究现状

国内外研究者对图像匹配开展了大量的研究工作，提出了很多的匹配方法。图像匹配主要包括：特征空间、搜索空间、搜索策略、相似性度量和决策策略共五个方面的问题。

基于图像灰度匹配和基于图像特征匹配是目前比较常见的匹配方法。蛋白质点匹配的本质就是在不同 2-DE 图像中寻找同源蛋白质点。在基于灰度的蛋白质点匹配方法中，根据 2-DE 图像上蛋白质点的大小和颜色深度等信息，直接对参

考图像和待匹配图像上的相关区域进行灰度相似性度量^[6]。根据相似性系数的大小,实现两 2-DE 图像间的蛋白质点匹配。而在基于特征的蛋白质点匹配方法中,主要依据参考图像和待匹配图像上蛋白质点对间的距离和角度变化关系等几何特征来实现蛋白质点匹配。

国外学者 Kham 和 Rahman 提出一种基于广义霍夫变换 (Generalized Hough Transform) 和 Canny 边缘检测的匹配算法^[7],但由于图像背景、蛋白质点大小及灰度等的变化,会造成较大的匹配误差;学者 Rogers 提出利用蛋白质点的中心坐标来实现匹配,在该方法中,将外形纹理 (Shap context) 和欧氏距离 (Euclidian Distance) 联合使用来估算最接近蛋白质点对的相似性^[8]。此外, Toru Wakahara 提出一种基于最邻近迭代匹配 (ICP) 算法^[9]; Grauman 提出一种基于图像金字塔的图像匹配算法^[10]。国内学者罗钟铉和刘成明提出了一种基于小波变换和投影特征的图像匹配的快速算法^[11],该算法大大提高了运算速度,同时又保持了匹配的精度,可以满足实时性的要求。

1.3 本课题的主要工作流程及本文的主要内容

图 1.2 显示了本课题的主要工作流程。

本文主要介绍了两幅二维凝胶电泳图像间的蛋白质点匹配问题,本文叙述到的算法实现都是基于 64 位 MATLAB7.13 版本实现的。文章内容如下:

第一章:绪论。在绪论中主要介绍了课题背景和蛋白质组学中的主要技术、研究意义以及当前蛋白质点匹配技术的研究现状。

第二章:基于 Landmark 的蛋白质点匹配算法。通过添加若干 Landmark 标记点,利用 `cp2tform` 函数推断变换参数,从而实现匹配。

第三章:基于 SURF 和 RANSAC 的蛋白质点匹配算法。利用 SURF 算法得到一组包含“异常点”的特征点对数据,然后调用 MATLAB7.13 工具箱函数完成 RANSAC 算法的迭代估计模型工作,利用得到的模型去除异常点,完成蛋白质点的匹配。

第四章:不同匹配算法的对比。在第四章中继续提出了一种基于几何相似性的匹配算法,即通过构造几何三角形进行蛋白质点匹配。接着对三种蛋白质点匹配算法进行了一个对比说明。

第五章:总结。主要是对本文算法的一个总结。

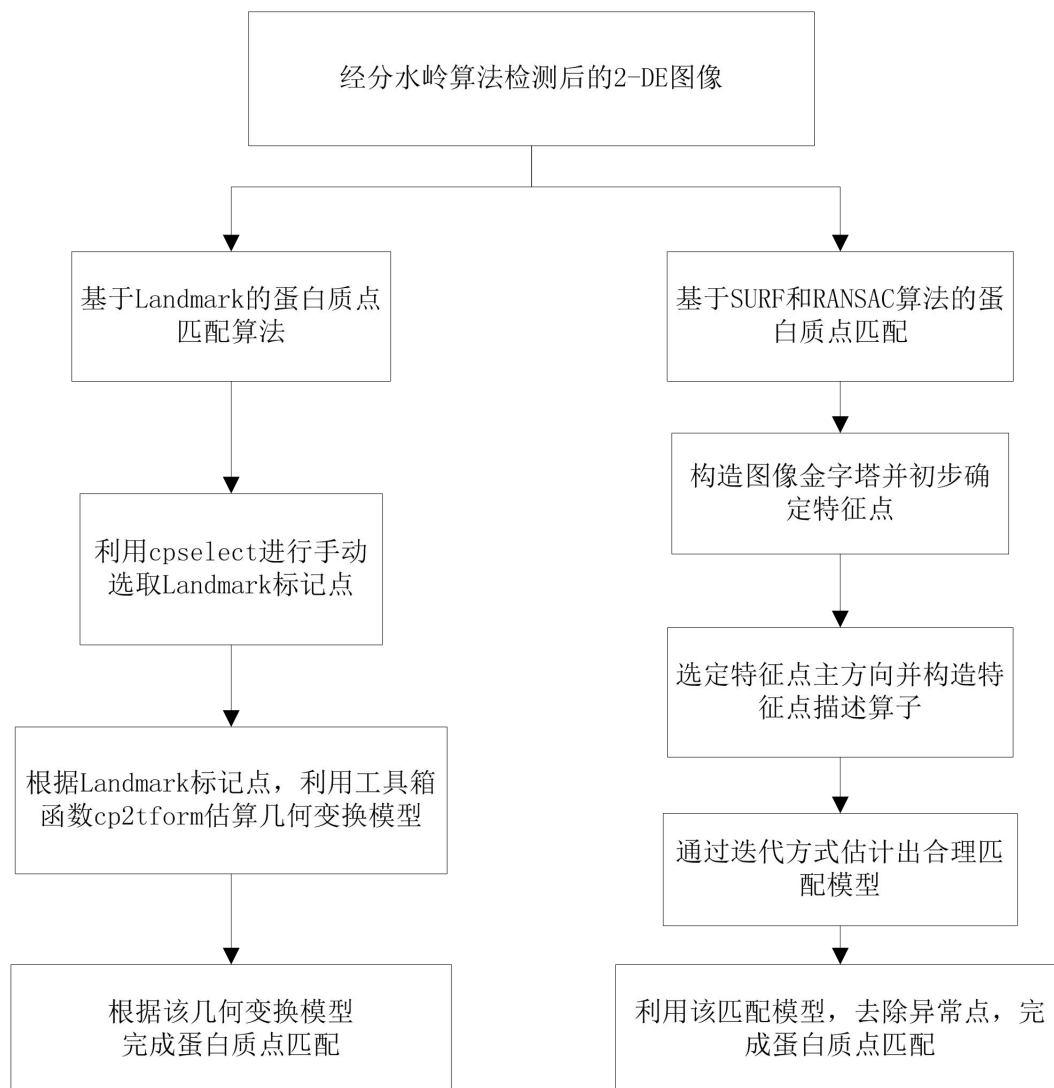


图 1.2 本课题的主要工作流程示意图

第二章 基于 Landmark 的蛋白质点匹配算法

蛋白质点匹配是在蛋白质点检测的基础上实现的。为便于算法叙述，参考 2-DE 图像和待匹配 2-DE 图像经分水岭算法检测后分别得到蛋白质点集 S 和 T 。其中， $S = \{p_i | i = 1, \dots, m\}$ ， $T = \{p_i | i = 1, \dots, n\}$ ， p 表示蛋白质点， m 和 n 分别表示参考 2-DE 图像和待匹配 2-DE 图像内的蛋白质点数。

2.1 概述

在获得 2-DE 图像的实验过程中，由于操作的原因，可能会造成蛋白质点的移动甚至是形变，而传统的 ICP 算法^[12]会造成误差累积，所以不适用于 2-DE 图像蛋白质点的精确匹配。在本章叙述的算法中，采用手动标记特征蛋白质点，然后选择合适的线性或非线性的几何变换模型，最终实现 2-DE 图像间的蛋白质点匹配。图 2.1 中给出了基于 Landmark 的蛋白质点匹配算法流程。

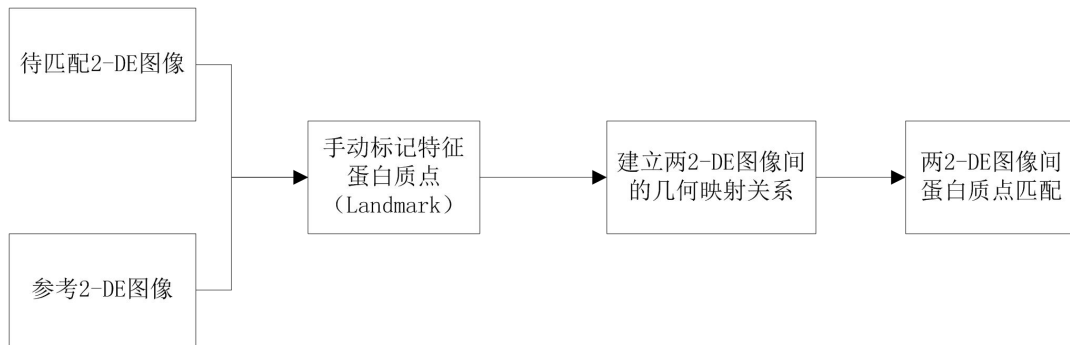


图 2.1 基于 Landmark 的蛋白质点匹配算法流程示意图

图 2.1 中各模块的主要过程为：

(1) 手动标记特征蛋白质点 (Landmark)：利用 `imread` 函数读取待匹配和参考 2-DE 图像，然后利用 `cpselect` 函数标记特征蛋白质点，并保存数据在 `.mat` 文件，MATLAB7.13 中 Landmark 标记界面如图 2.2 所示，其中左右图分别表示经分水岭算法检测蛋白点后得到的待匹配 2-DE 图像和参考 2-DE 图像。

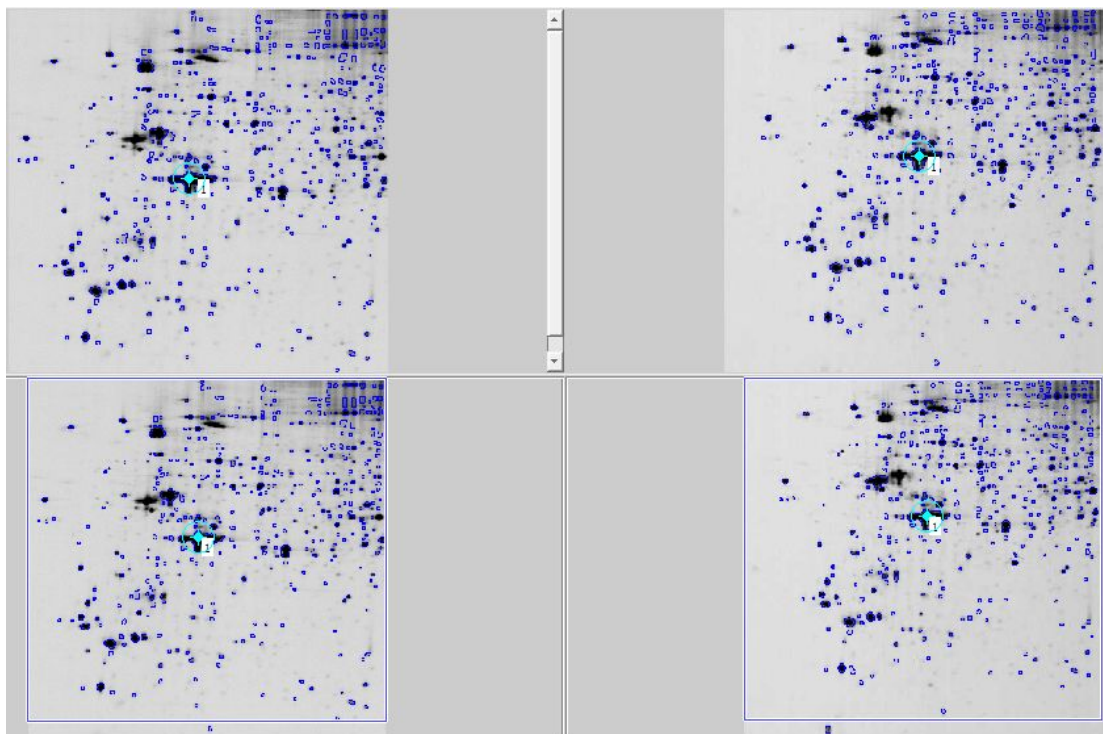


图 2.2 MATLAB 中 Landmark 标记点选取界面

(2) 建立两图像间的几何映射关系：根据 (1) 中选取的特征蛋白质点，采用合适的几何变换模型，并建立映射关系。

(3) 两 2-DE 图像间蛋白质点匹配：根据几何映射关系，对两幅 2-DE 图像中的蛋白质点进行映射，并建立统一的坐标系，实现 2-DE 图像间的蛋白质点匹配。

2.2 Landmark 标记点的选取

本章的蛋白质点匹配的前提条件是 Landmark 标记，Landmark 标记点是采用手动选取方式获得的，对两 2-DE 图像中特征明显的对应蛋白质点进行标记，并把这些标记点作为蛋白质点匹配的控制点。如果 Landmark 标记不合理则会造成后续蛋白质点的误匹配，还有，如果 Landmark 标记点选取的过少，则不能真实的反映出 2-DE 图像间的映射关系；同样，过多则需要较多的人工，而且计算量较大。因此，Landmark 标记一般遵从以下规则：

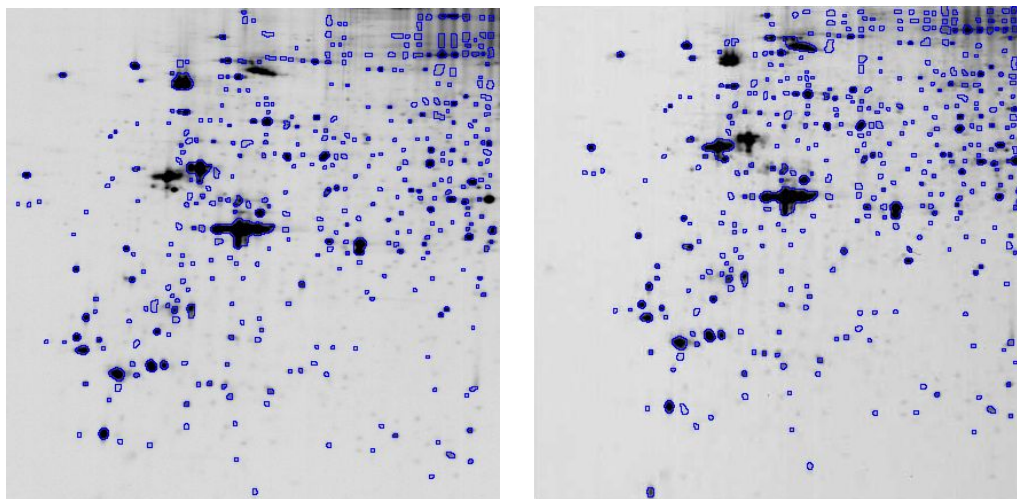
(1) 选取具有鲜明特征的蛋白质点作为 Landmark 标记点。如果选取的蛋白质点的灰度与整幅图像的背景灰度相似，即很难通过颜色来判断是否为蛋白质点，则不能作为 Landmark 标记点；针对这个问题，在本章选取特征蛋白质点的过程中，采用经过分水岭算法处理过的 2-DE 图像（如图 2.3，其中分水岭线内为蛋白质点）作为载体，在其上进行标记。此外，如果蛋白质点因为某些不可抗

因素只在其中一副 2-DE 图像中存在或消失，则这类点亦不能够选为 Landmark 标记点。

(2) 尽量选取坐标位置一致的蛋白质点作为 Landmark 标记点。由于 2-DE 图像是在不同的条件下获得的，即使是同一蛋白质组织的 2-DE 图像上的蛋白质点也会有一定的形变，从而在蛋白质点检测中，对两 2-DE 图像中的同一蛋白质点的坐标位置计算存在较大误差。这类点不能作为 Landmark 标记点，如图 2.4 中所示，同一蛋白质点，其坐标为（以“+”标记）在参考 2-DE 图像中在该蛋白质点的上部位置，而在待匹配 2-DE 图像中则是位于靠下位置，这类的蛋白质不能选作特征点。

(3) Landmark 标记点应均匀分布于 2-DE 图像。如果 Landmark 标记蛋白质点选取不均匀，则会建立存在较大误差的变换模型，从而导致错误的蛋白质点匹配。因此，均匀选取 Landmark 标记点是必要的，而且蛋白质点较为密集的区域标记特征蛋白质点的数目也应越多。

(4) 合理选取 Landmark 标记点的数量。根据选取的图像变换模型，在满足该变换模型最小量的基础上，应根据 2-DE 图像的形变程度和蛋白质点的总量选取合适的 Landmark 标记点数，这样既能保证模型的不致失真，又可以把工作量降到最低，以提高效率。



(a) 参考 2-DE 图像

(b) 待匹配 2-DE 图像

图 2.3 2-DE 图像经分水岭算法处理后的图像



(a) 坐标在参考 2-DE 图像中的位置

(b) 坐标在待匹配 2-DE 图像中的位置

图 2.4 同一蛋白质点的坐标在两图像中的差异

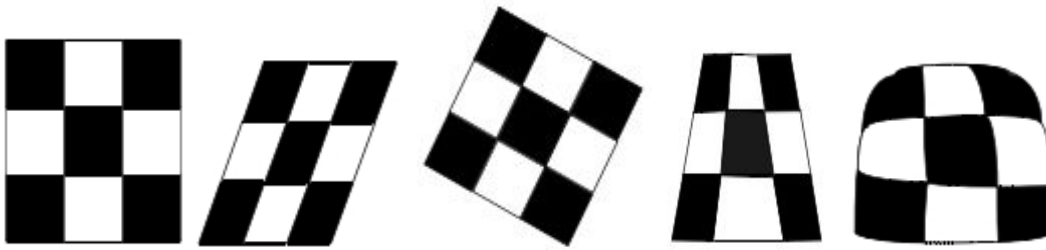
2.3 选取两凝胶图像间的几何映射模型

选取合适的几何变换模型，是能否匹配成功的关键因素。在本章算法中通过调用 MATLAB 图像处理工具箱中的 `cp2tform` 函数^[13]来推断几何变换参数。`cp2tform` 的语法是：

`tform=cp2tform(input_points,base_points,transformtype)`

参数 `input_points` 和 `base_points` 是前一节中选取的 Landmark 标记点集，第三个参数 `transformtype` 是指定的所希望的几何变换类型。在本章的蛋白质点匹配算法中，采用的是透视变换。

图 2.5 给出了几种常见的几何变换模型。这些空间变换的变换矩阵和适用性分别如下所述：



(a) 原始图像

(b) 仿射变换

(c) 刚体变换

(d) 透视变换

(e) 非线性变换

图 2.5 图像空间变换模型示意图

(1) 仿射变换 (Affine)

仿射变换不仅存在镜像、旋转和平移变换，还包括尺度变换和剪切变换等变换。在二维图像中，点 (u,v) 仿射变换为点 (x,y) 的变换公式为：

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} m_1 & m_2 \\ m_4 & m_5 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} + \begin{bmatrix} m_3 \\ m_6 \end{bmatrix} \quad (2-1)$$

其变换矩阵 A 可记为：

$$A = \begin{bmatrix} m_1 & m_2 & m_3 \\ m_4 & m_5 & m_6 \\ 0 & 0 & 1 \end{bmatrix} \quad (2-2)$$

(2) 刚体变换 (Nonreflective similarity)

刚体变换仅存在镜像、旋转和平移变换。点(u,v)刚体变换为点(x,y)的变换公式为:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \end{bmatrix} \quad (2-3)$$

其中, θ 表示图像的旋转角度, n_1 和 n_2 分别表示图像在横向和纵向上的平移量。其变换矩阵 N 可记为:

$$N = \begin{bmatrix} \cos \theta & -\sin \theta & n_1 \\ \sin \theta & \cos \theta & n_2 \\ 0 & 0 & 1 \end{bmatrix} \quad (2-4)$$

(3) 透视变换 (Projective)

透视变换又称为投影变换,是由成像角度的改变而造成的。点(u,v)透视变换为点(x,y)的变换公式为:

$$\begin{cases} x = \frac{m_1 u + m_2 v + m_3}{m_7 u + m_8 v + 1} \\ y = \frac{m_4 u + m_5 v + m_6}{m_7 u + m_8 v + 1} \end{cases} \quad (2-5)$$

其变换矩阵 P 可记为:

$$P = \begin{bmatrix} m_1 & m_2 & m_3 \\ m_4 & m_5 & m_6 \\ m_7 & m_8 & 1 \end{bmatrix} \quad (2-6)$$

(4) 非线性变换 (Nonlinear transformation)

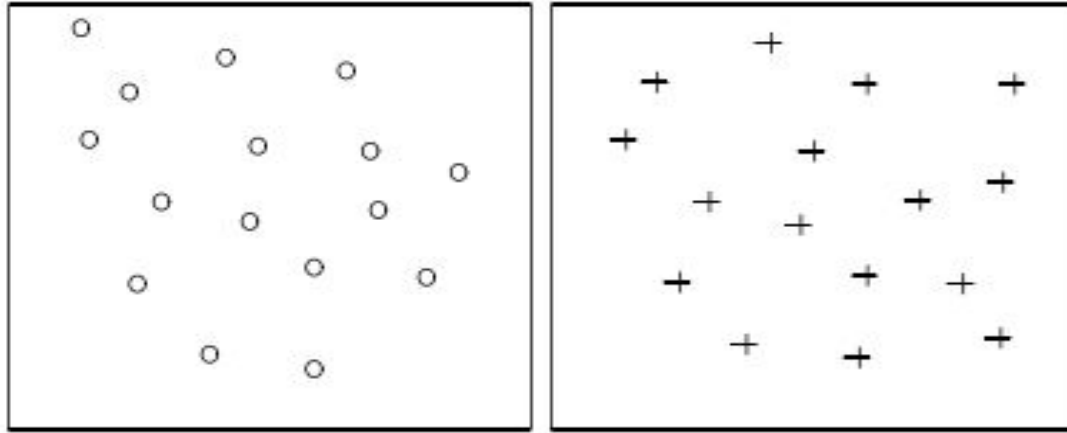
非线性变换常应用于具有全局非线性形变或整体近似刚体变换而具有局部非线性形变的图像匹配过程中。点(u,v)非线性为点(x,y)的变换公式为:

$$\begin{cases} x = f_1(u, v) \\ y = f_2(u, v) \end{cases} \quad (2-7)$$

2.4 两 2-DE 图像间的蛋白质点匹配

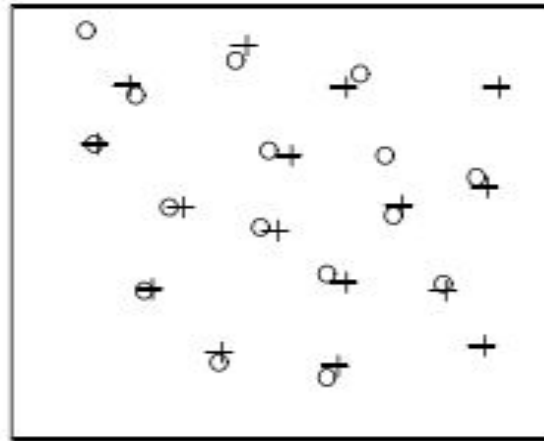
蛋白质点匹配是将不同坐标系下的蛋白质点变换到同一坐标系下完成的。如

图 2.6 所示，参考 2-DE 图像中的蛋白质点用“o”表示，待匹配 2-DE 图像中的蛋白质点用“+”表示。



(a) 参考 2-DE 图像中的蛋白质点

(b) 待匹配 2-DE 图像中的蛋白质点



(c) 变换到同一坐标系内的蛋白质点

图 2.6 两 2-DE 图像中的蛋白质点变换到同一坐标系内

坐标变换完成后，采用最小距离度量法去完成蛋白质点对的匹配。在本算法中，这些工作均由 `cp2tform` 函数自动处理。对于参考 2-DE 图像中蛋白质点 A 和待匹配 2-DE 图像中的蛋白质点 B，A 和 B 若为同源蛋白质点，那么在蛋白质点 A 和 B 经坐标变换到同一坐标系后，应满足如下关系：

(1) 在所有参考 2-DE 图像的蛋白质点中，B 与 A 之间的欧式距离 (d_{AB}) 应是最小的；

(2) $d_{AB} \leq d_r$ ，距离度量阈值 d_r 一般取 10 个单位像素点宽度。

图 2.7 中，给出了基于 Landmark 的蛋白质点匹配算法的实现流程。

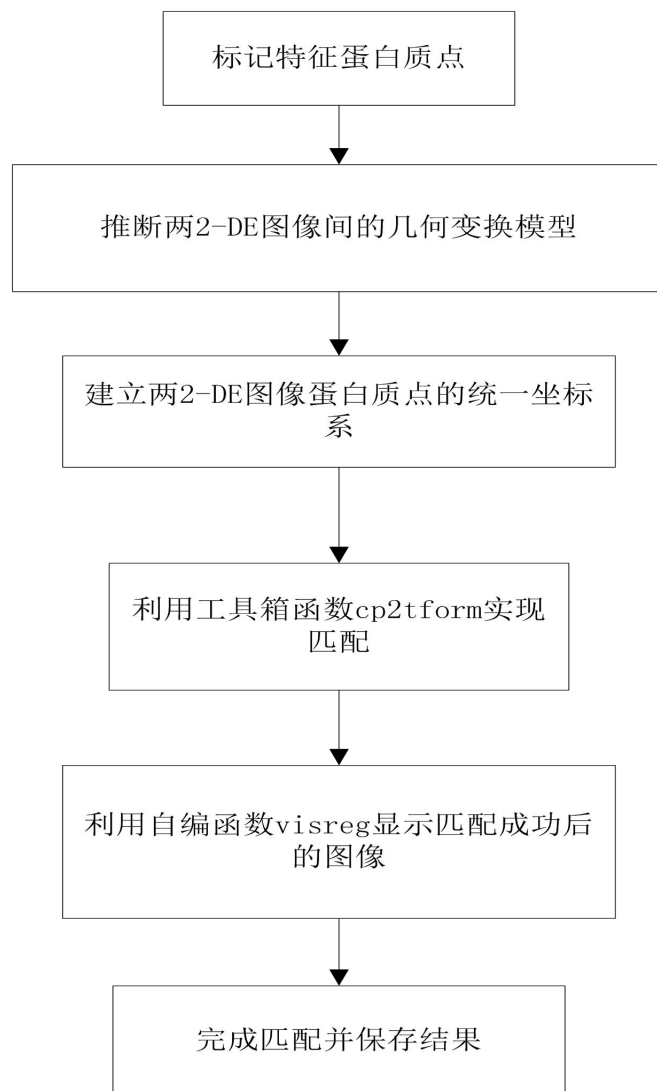
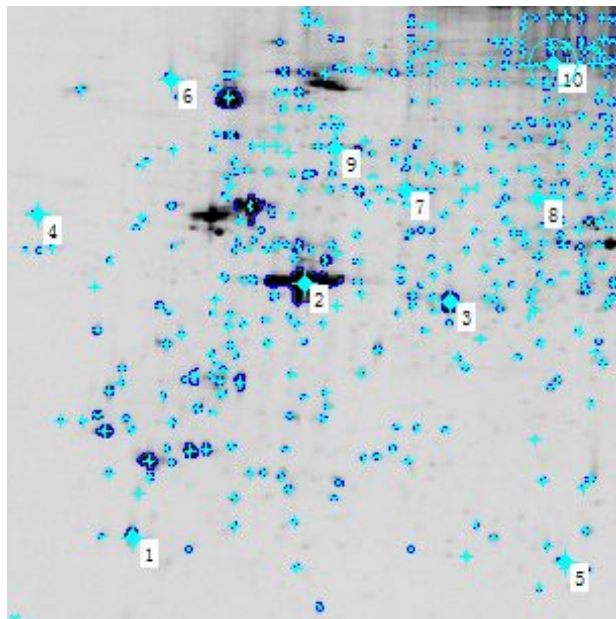


图 2.7 基于 Landmark 的蛋白质点匹配算法流程

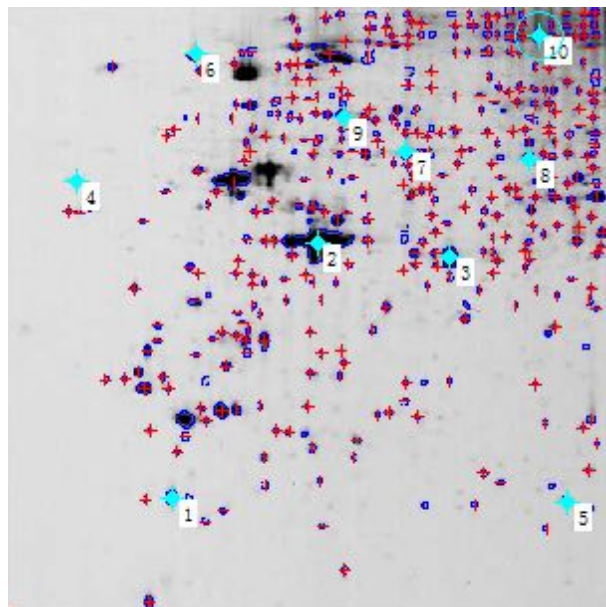
2.5 实验结果及分析

为了验证本章算法的可行性，分别采用不同的标记点数目进行实验验证，分别标记 10、30、42 个 Landmark 标记点。为了使匹配效果看的更明显，参考 2-DE 图像中的蛋白质点中心用“*”表示，待匹配 2-DE 图像中的蛋白质点中心用“+”表示。实验结果如下：

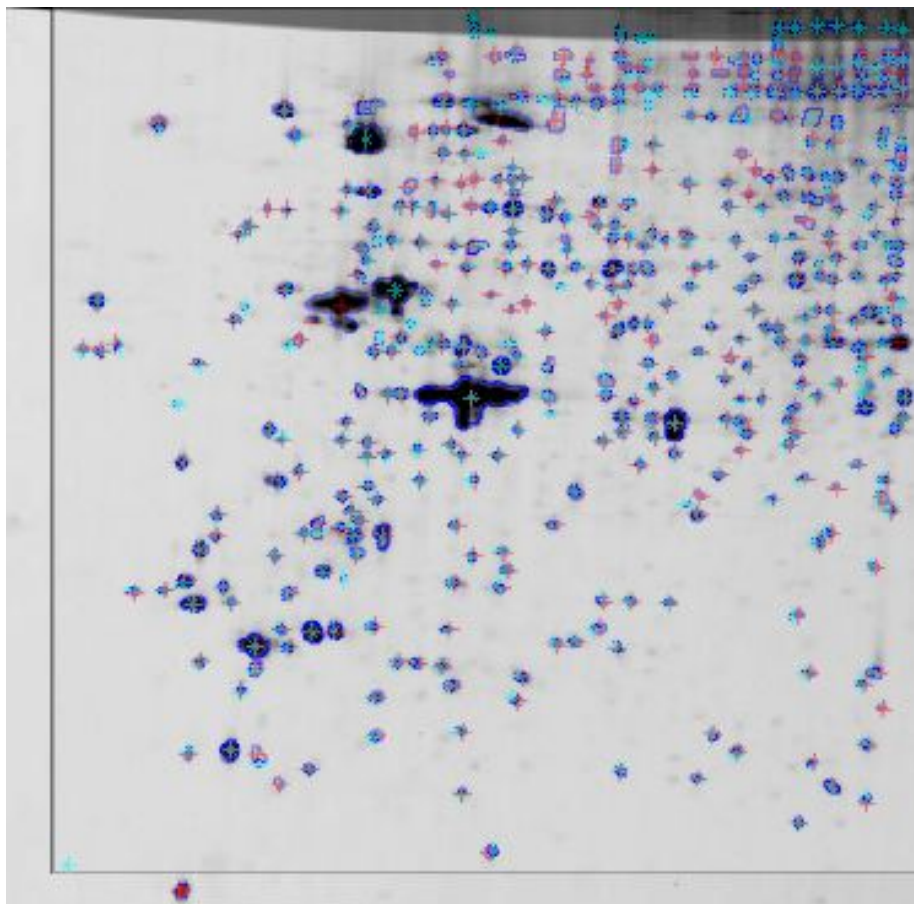
- (1) 当 Landmark 标记点数为 10 时，实验结果如图 2.8 所示。图 2.8 (a)、(b) 分别是参考图像和待匹配图像，其中标记点用“1、2、3...”显示。



(a) 参考 2-DE 图像



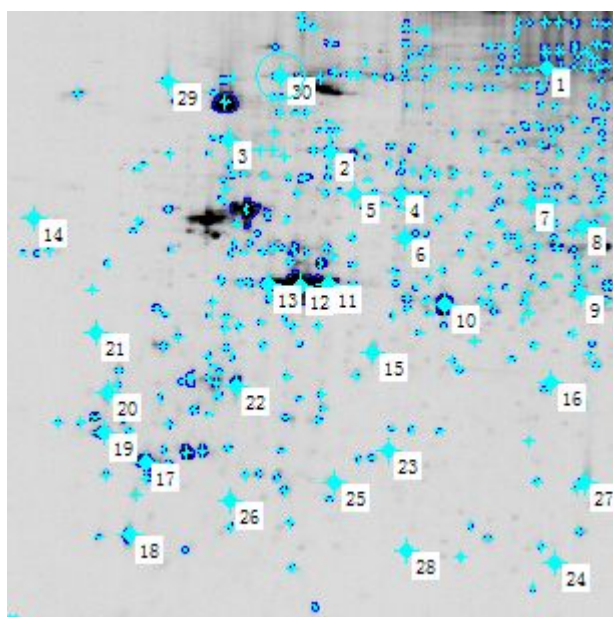
(b) 待匹配 2-DE 图像



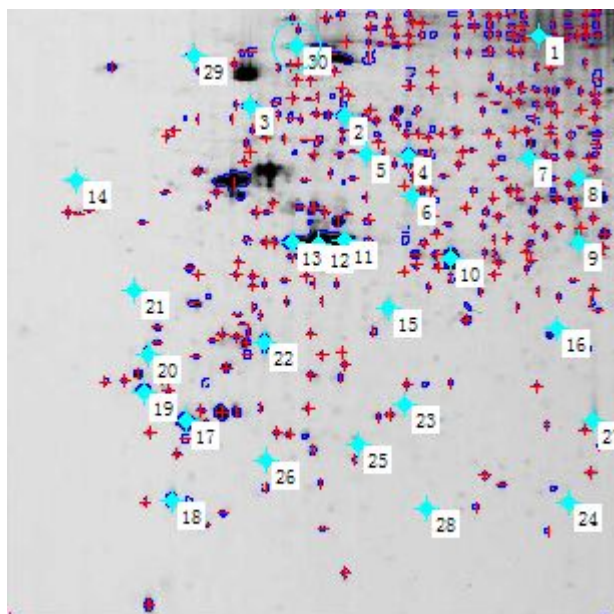
(c) 匹配结果

图 2.8 Landmark 标记点数为 10 时的匹配情况

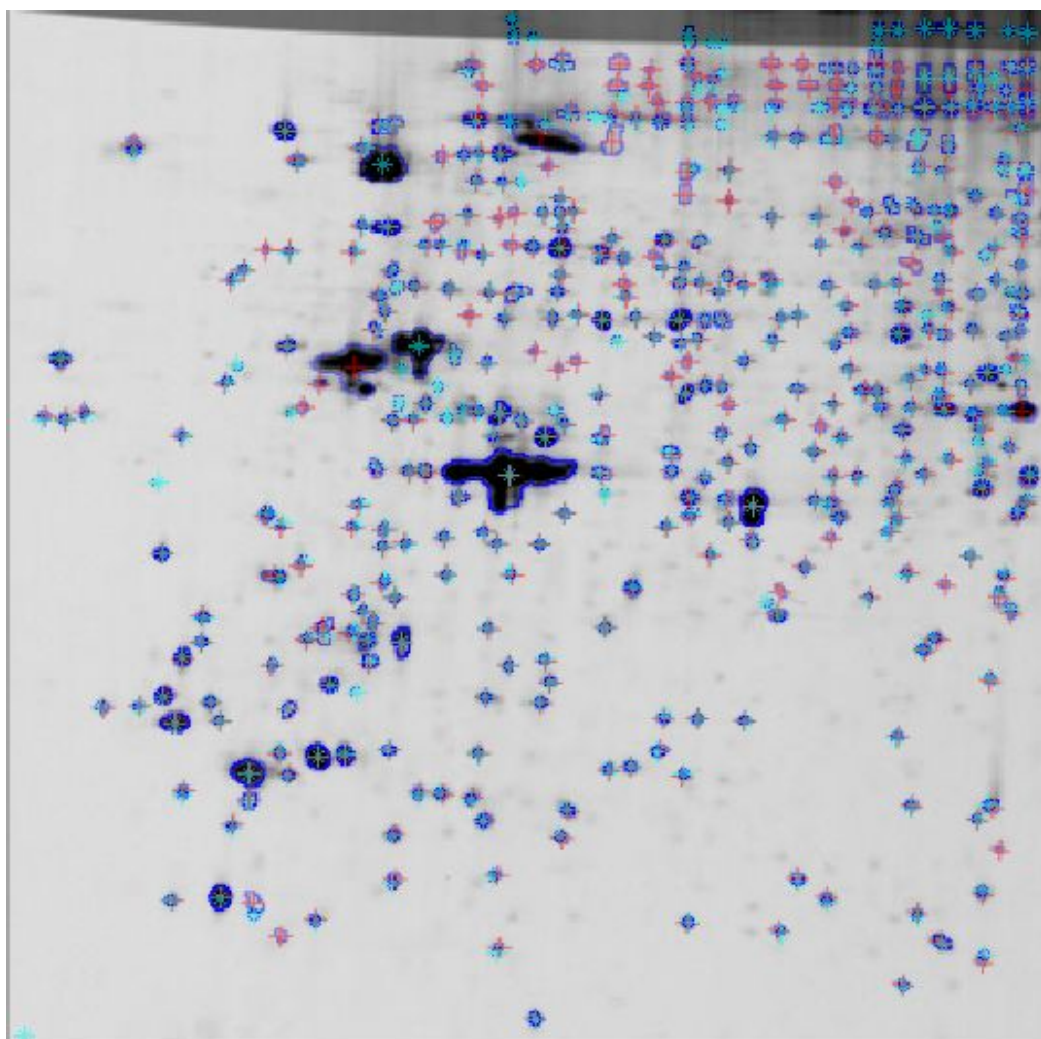
(2) 当 Landmark 标记点数为 30 时，实验结果如图 2.9 所示。



(a) 参考 2-DE 图像



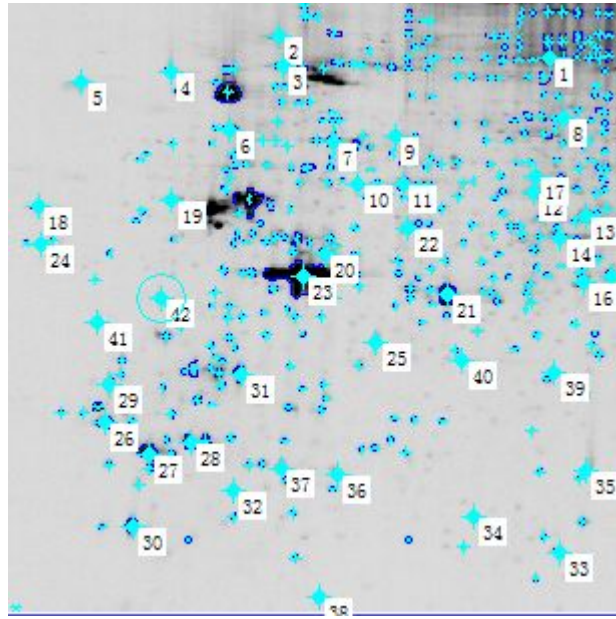
(b) 待匹配 2-DE 图像



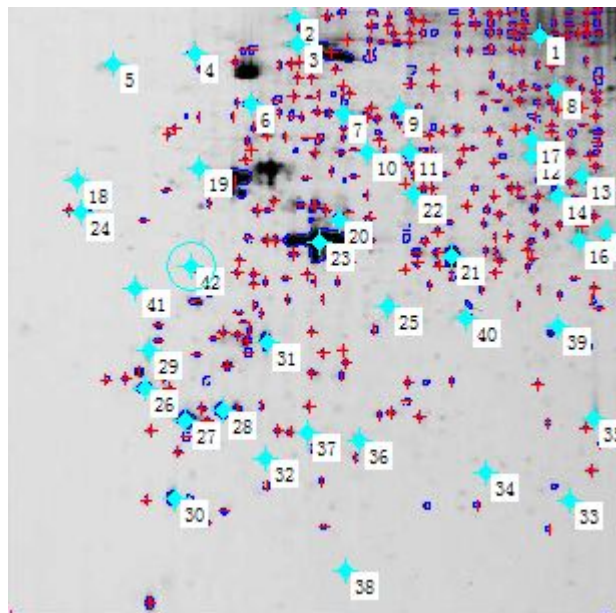
(c) 匹配结果

图 2.9 Landmark 标记点数为 30 时的匹配情况

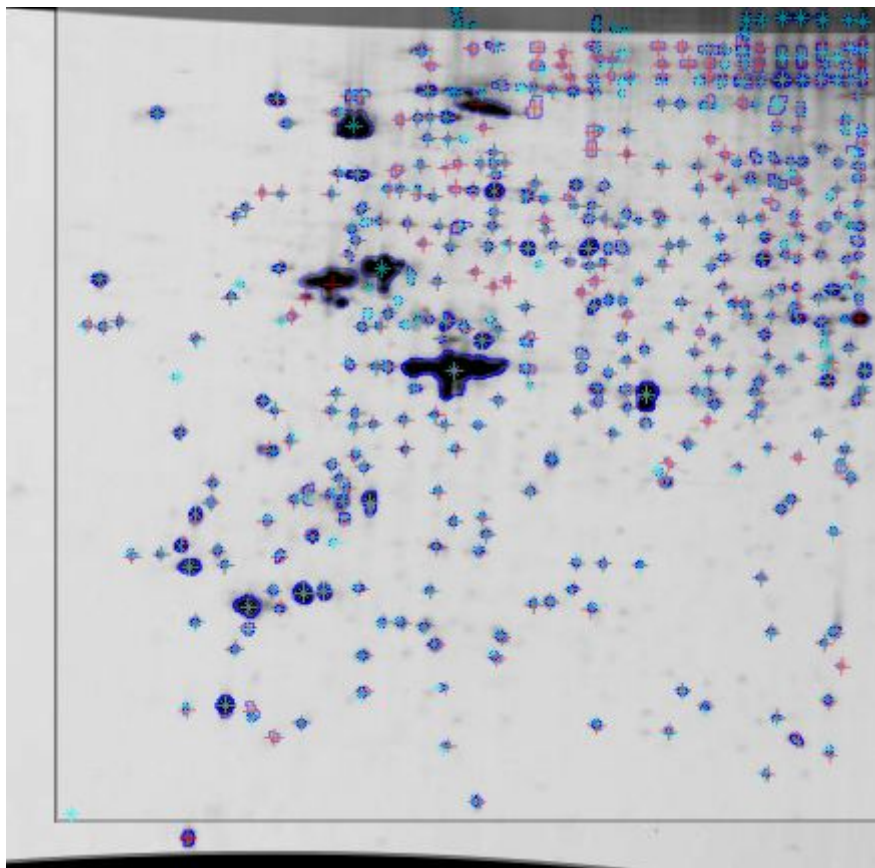
(3) 当 Landmark 标记点数为 42 时，实验结果如图 2.10 所示。



(a) 参考 2-DE 图像



(b) 待匹配 2-DE 图像



(c) 匹配结果

图 2.10 Landmark 标记点数为 42 时的匹配情况

(4) 表 2.1 给出了 (1)、(2)、(3) 中三种不同 Landmark 标记点下完成的匹配点对个数，其中匹配率 $E = \frac{2 \times U}{B + A} \times 100\%$ 。

表 2.1 三种不同 Landmark 标记点数下的匹配结果

Landmark 个数	蛋白点检测数目		本算法匹配 的点对数	匹配率 E (%)
	参考图像个数B	待匹配图像个数A		
10	392	393	325	82.8
30	392	393	338	86.1
42	392	393	341	86.7

由表 2.1 中的实验结果可以看出，本章叙述的算法可以完成一定数量的蛋白质点匹配。可以发现，在本章算法下，当 Landmark 标记点数越多时，匹配的效果越好。从匹配效果图中可以发现，蛋白质点误匹配区域主要出现在蛋白质点形变或移动区域，由于该区域一般处于 2-DE 图像的边缘，而且蛋白质点边缘不是很明显，在检测蛋白质点的时候会存在一定的误差，最终累积误差造成蛋白质点未能匹配。此外，本章叙述的匹配算法虽然通过手动标记特征蛋白质点确定

2-DE 图像间的几何变换模型，减少了图像形变对蛋白质点匹配的误差影响；可是，由于在本算法中采用的全局变换模型，即以整张 2-DE 图像作为变换的本体，然而这样做忽略了 2-DE 图像中局部形变的情况，所以这种算法难以满足高精度匹配的要求。

2.6 本章小结

在本章叙述的算法中，采用手动添加 Landmark 标记点标记特征蛋白质点的方法，对检测到的蛋白质点进行特征提取；然后利用 MATLAB 工具箱中的 cp2tform 函数去推断变换参数，并确立 Landmark 标记点间的几何变换模型，从而利用这个模型对两 2-DE 图像建立统一的空间坐标系，进行 2-DE 图像间的蛋白质点匹配。通过比较不同数量的 Landmark 标记点来检验本算法的匹配效果，从匹配结果图可以发现，当 Landmark 标记数越多而且在凝胶图像上分布越均匀的时候，匹配效果越好；说明本章算法具有一定的可行性，能够实现一定精度要求下的 2-DE 图像蛋白质点匹配。当然，通过实验结果可以看出本算法的局限性也是比较大的，主要体现在，对部分存在干扰的邻近相似蛋白质点匹配和局部形变的蛋白质点进行匹配时，会发生误差匹配，而且会漏掉一些蛋白质点未能匹配，导致 2-DE 图像蛋白质点匹配的正确率较低。此外，本算法的人工介入较多，降低了蛋白质点匹配的自动化程度，而且手动添加 Landmark 标记点时也存在一定的人为误差。

第三章 基于 SURF 和 RANSAC 的蛋白质点匹配算法

由第二章的匹配结果可以看出，基于 Landmark 的蛋白质点匹配算法的人工介入较多，在匹配大量的 2-DE 图像时是不适用的，故在本章提出一种基于 SURF 和 RANSAC 的 2-DE 图像蛋白质点自动匹配算法。

3.1 概述

3.1.1 SURF 算法

1、SURF 算法的基本概念：

SURF 是“Speed Up Robust Features（快速鲁棒特征）”的缩写，是一种尺度和旋转不变的兴趣点 detector（检测器）和 descriptor（描述符），是尺度不变特征变换算法（SIFT 算法^[14]）的加速版。速度快是该算法的最大特点。SURF 采用的主要策略有：1）积分图像，用于卷积图像；2）detector 是基于 Hessian 矩阵，descriptor 是基于分布的。

（1）积分图像

积分图像计算的是某一个矩形区域内的像素和，积分图像有助于加速算法。

积分图 $I_{\Sigma}(x)$ 在位置 $x = (x, y)^T$ 处的定义为：

$$I_{\Sigma}(x) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j) \quad (3-1)$$

计算一幅图片中的任意一个矩形区域内像素点的求和 Σ 运算只需要三个加（减）法运算即可（如图 3.1 所示），而且算子 Σ 独立于矩形区域的大小。

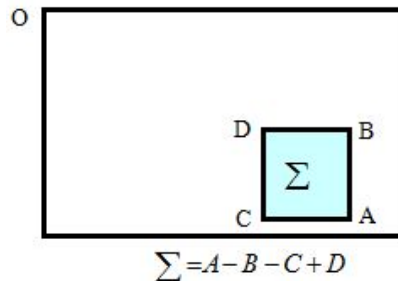


图 3.1 图像中某个矩形区域内的像素点求和

（2）Hessian 矩阵

SURF 对于兴趣点的检测是基于最基本的 Hessian 近似矩阵完成的。给定图

像 I 中的一点 $X = (x, y)$ ，Hessian 矩阵 $H(X, \sigma)$ 在 X 处尺度为 σ 的定义为：

$$H(X, \sigma) = \begin{bmatrix} L_{xx}(X, \sigma) & L_{xy}(X, \sigma) \\ L_{xy}(X, \sigma) & L_{yy}(X, \sigma) \end{bmatrix} \quad (3-2)$$

这里， $L_{xx}(X, \sigma)$ 是高斯二阶偏导数 $\frac{\partial^2}{\partial x^2}$ 在 X 处图像 I 的卷积。

2、SURF 算法提取特征点实现步骤：

(1) 构造图像多尺度空间

算法的尺度不变性主要是在不同尺度下寻找兴趣点。Lowe 在其 SIFT 算法中这样构造尺度空间：对原图像不断的进行高斯平滑+降采样，得到金字塔图像后，又进一步得到 Dog 图像，边和斑状结构就是通过 Dog 图像得到其在原图中的位置。SURF 中略有不同的是，SIFT 算法在构造金字塔图层的时候保持高斯滤波器大小不变，改变的是图像的大小；而在 SURF 算法中的是保持图像的大小不变，改变的是高斯滤波器的大小。如图 3.2 所示，图中左为 SIFT，右为 SURF。

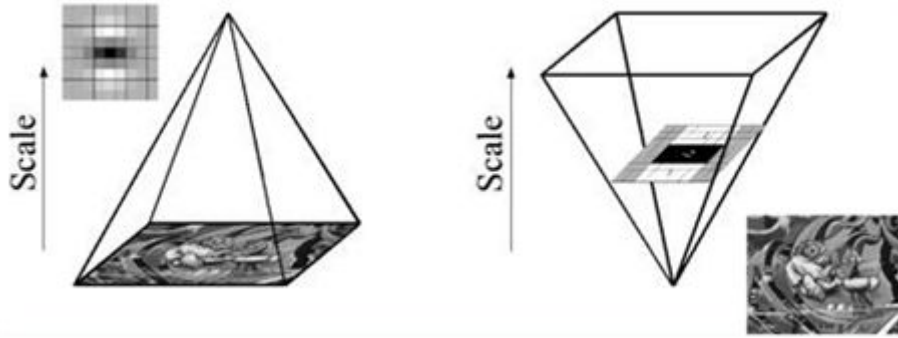


图 3.2 SIFT 和 SURF 算法构造尺度空间对比

从图 3.2 可以看出：假设初始模板的大小为 9×9 ，尺度 $scale$ 定为 1.2，对应于高斯核中的 $\sigma = 1.2$ （这是第一层），然后不断增加模板的尺寸，得到下一层。这样做可以提高效率，而且，因为没有对图像进行降采样，所以不存在混叠现象。与 SIFT 类似，SURF 的尺度空间也是按组划分的，每一个组里对输入图像用尺度不断增加的滤波器进行滤波后得到一系列响应。总体上，一组包含了一个缩放因子，每一组内的层数是一常数。

(2) 提取特征点

所以小于预设极值的取值都被丢弃，增加预设极值使检测到的特征点数量减少，最终只有几个特征最强点会被检测出来。检测过程中使用与该尺度层图像解析度相对应大小的滤波器进行检测，以 3×3 的滤波器为例（如图 3.3 所示），中间的检测点和它同尺度的 8 个相邻点和上下相邻尺度对应的 9×2 个点，共 26 个

点比较，图中标记‘X’的像素点的特征值若大于周围像素则可确定此点为此区域的特征点。

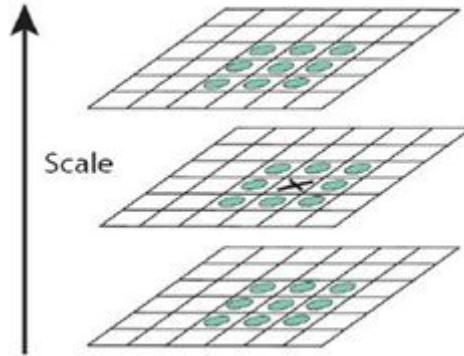


图 3.3 3×3 滤波器示意图

(3) 选取特征点主方向

如图 3.4 所示，为保证旋转不变性，首先以特征点为中心，计算半径为 $6s$ (s 为特征点所在的尺度值) 的邻域内的点在 z 、 y 方向的 HARR 小波 (HARR 小波边长取 $4s$) 响应，并给这些响应值赋高斯权重系数，使得靠近特征点的响应贡献大，而远离特征点的响应贡献小。其次将 $6s$ 范围内的响应相加以形成新的矢量，遍历整个圆形区域，选择最长矢量的方向为此特征点的主方向。通过对特征点逐个进行计算，即可得到所有特征点的主方向。

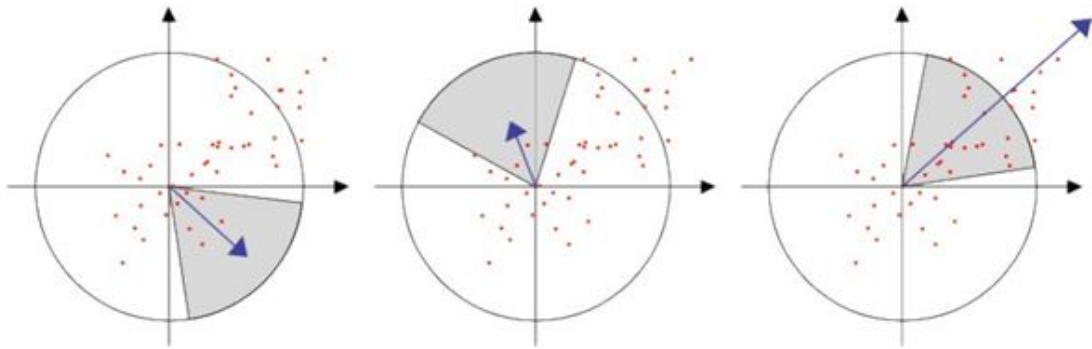


图 3.4 选取特征点主方向示意图

(4) 构造特征点算术描述子

先将坐标轴旋转为关键点方向，确保旋转不变性。为了增加鲁棒性，可对 dx, dy 进行高斯滤波 ($\sigma = 3.3s$)，滤波器的中心点为兴趣点。在图 3.5 中，右边的小框即是实际中 descriptor 的基本单元。对于每一个子域 (图 3.5 中右侧所示)，分别计算四个量，即 $\sum dx$ ， $\sum |dx|$ ， $\sum dy$ ， $\sum |dy|$ 。现可用矢量 $v = (\sum dx, \sum |dx|, \sum dy, \sum |dy|)$ 来表示每一个子区域。将 16 个矢量连在一起，即得到一个长为 64 的 descriptor 矢量，用于描述大块-区域。如果对其进行归一化，

得到的矢量则具有尺度不变性。通过图 3.6 可以看出，采用四个量描述区域可以更加具有区分度。

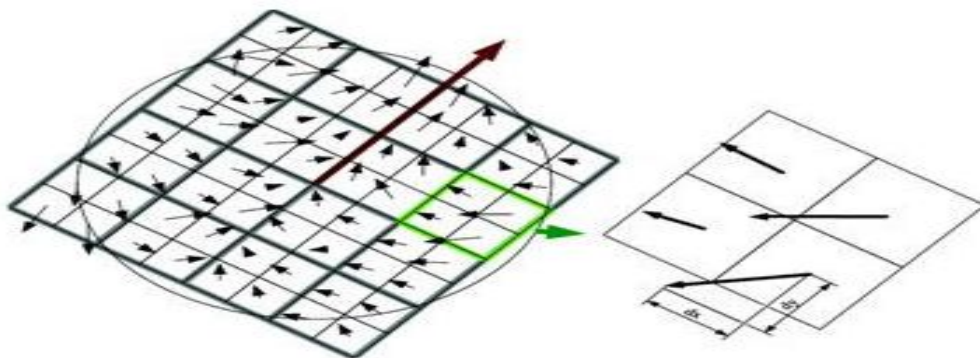


图 3.5 构造算术描述子示意图

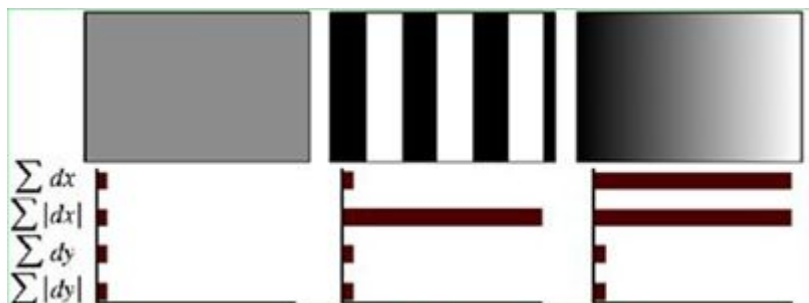


图 3.6 四个量描述不同的区域示意图

3.1.2 RANSAC 算法

1、RANSAC 算法的基本概念：

RANSAC 是“Random Sample Consensus (随机抽样一致)”的缩写。此算法可以从一组包含“异常点”的观测数据集中，通过迭代方式估计数学模型的参数。它是一种不确定的算法，即它有一定的概率得出一个合理的结果。为了提高合理结果的概率，就必须增加迭代的次数。该算法最早由 Fischler 和 Bolles^[15]在 1981 年提出。在 RANSAC 提出了三种基本假设，即：

- (1) 数据由“局内点”组成，数据的分布可以用模型参数来解释；
- (2) “异常点”数据是不能适应该模型的，“异常点”一般是由噪声的极值、错误的测量方法以及对数据的错误假设造成的；
- (3) 除此之外的数据都属于噪声。

在图 3.7 中给出了利用 RANSAC 算法找出一条合适的二维直线的示例。现假设观测数据中包含局内点和异常点，其中局内点近似附在一条直线上，而异常点在直线之外。因为最小二乘法会尽量去拟合所有点，所以无法得出合适的直线。

然而 RANSAC 算法有很高的概率仅用局内点就可得出一个合适的直线模型，但是 RANSAC 算法不能确定结果完全无误。所以此时，选择合适的参数是必须的。

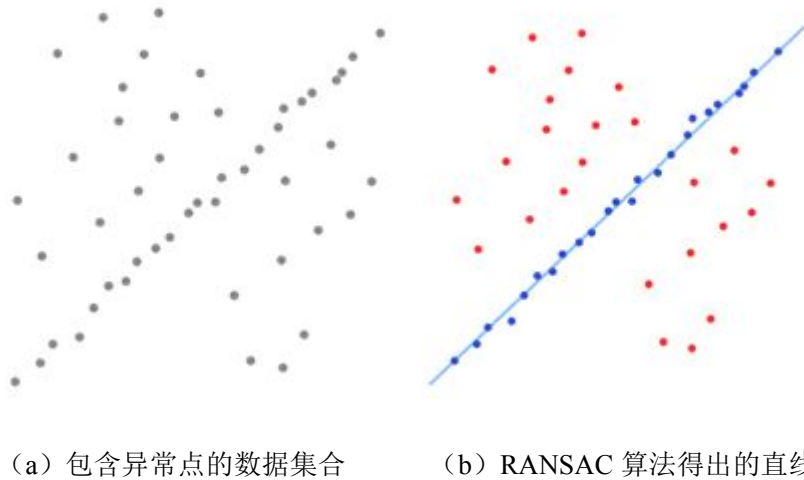


图 3.7 RANSAC 算法拟合直线示意图

RANSAC 算法输入内容有观测数据集、适应观测数据集的参数化模型以及可信的参数。RANSAC 算法通过不断地选择观测数据集中的一组随机子集来实现，其中被选中的子集被假设为局内点，通过以下步骤实现对局内点的验证：

- (1) 有一个模型适合假设的局内点，使得所有未知参数都能利用假设的局内点得出；
- (2) 用 (1) 中得到的模型监测其它数据，若有点适合估计的模型，即认为此点为局内点；
- (3) 若足够多的点被认为是局内点，则估计的模型就是合理的；
- (4) 将所有假设的局内点进行重新估计模型；
- (5) 用局内点与模型的错误率来评估模型的合理性。

此过程会被重复执行一定的次数，每次产生的模型要么因包含局内点过少而被弃用，要么比当前模型更合理而被选用。

2、确定参数：

对于特定的问题及数据集需要通过实验来确定参数 t （用于决定数据是否适应于模型的阈值）和参数 d （用于判定模型是否适应于数据集的数据数目），然而参数 k （算法的迭代次数）可从理论得出。现用 p 表示迭代过程中从数据集内随机取出的点都是局内点的概率； w =局内点数目/数据集的数目，表示从数据集选出局内点的概率；假设估计模型需要选定 n 个点， w^n 是所有 n 个点都是局内点的概率；用 $(1-w^n)^k$ 来表示算法永远选不到 n 个点都是局内点的概率，则有

$1-p=(1-w^n)^k$ ，则

$$k = \frac{\log(1-p)}{\log(1-w^n)} \quad (3-3)$$

这是假设 n 个点都是独立选择得到的结果，那么在某个点被选中之后，可能会在后面的迭代过程中被重复选到，而此时推导出的 k 值表示选取不重复点的上限，这是不合理的。为了得到更可信的参数，通常将标准偏差或其乘积添加到 k 上。 k 的标准差为：

$$SD(k) = \frac{\sqrt{(1-w^n)}}{w^n} \quad (3-4)$$

3.2 2-DE 图像间蛋白质点匹配

图 3.8 给出了本章算法进行蛋白质点匹配的流程。

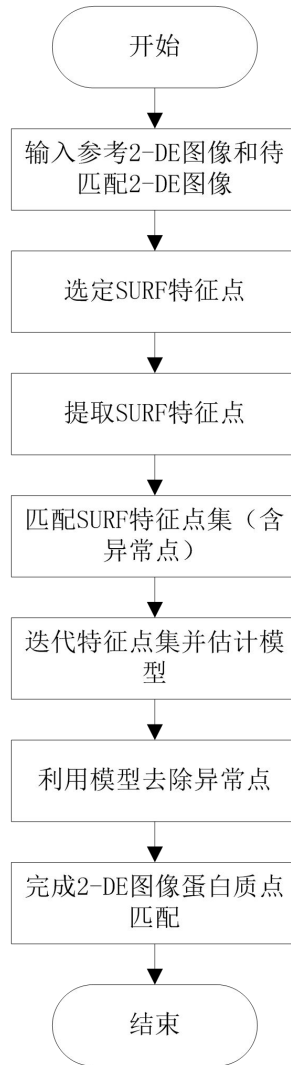


图 3.8 基于 SURF 和 RANSAC 的蛋白质点匹配算法流程图

在本章叙述的算法中，利用了 MATLAB7.13 版本的工具箱函数 `detectSURFFeatures` 完成 SURF 特征点的选定工作；接着利用 `extractFeatures` 函数完成对 SURF 特征点的提取工作；再用 `matchFeatures` 函数对提取到的特征点进行特征匹配，接着就可以得到包含“outliers 点”（即包含“异常点”）的匹配特征点集。通过对包含异常点的匹配特征点集进行迭代处理，并构造一个合理的变换模型，利用这个合理的变换模型去除异常点，从而完成对 2-DE 图像蛋白质点的匹配。所有在 3.1 节算法叙述中涉及到的参数计算，全部由 MATLAB 工具箱函数进行自动处理。本章的主要代码部分在附录中给出。

3.3 实验结果及分析

本章算法对 2-DE 图像蛋白质点匹配的效果如图 3.9 所示，用向量的形式表示蛋白质点对间的匹配。图 3.9（a）显示的是包含“异常点”的 2-DE 图像匹配结果，图 3.9（b）显示的是去除异常点之后的匹配结果。



图 3.9（a）包含“异常点”的蛋白质点匹配结果

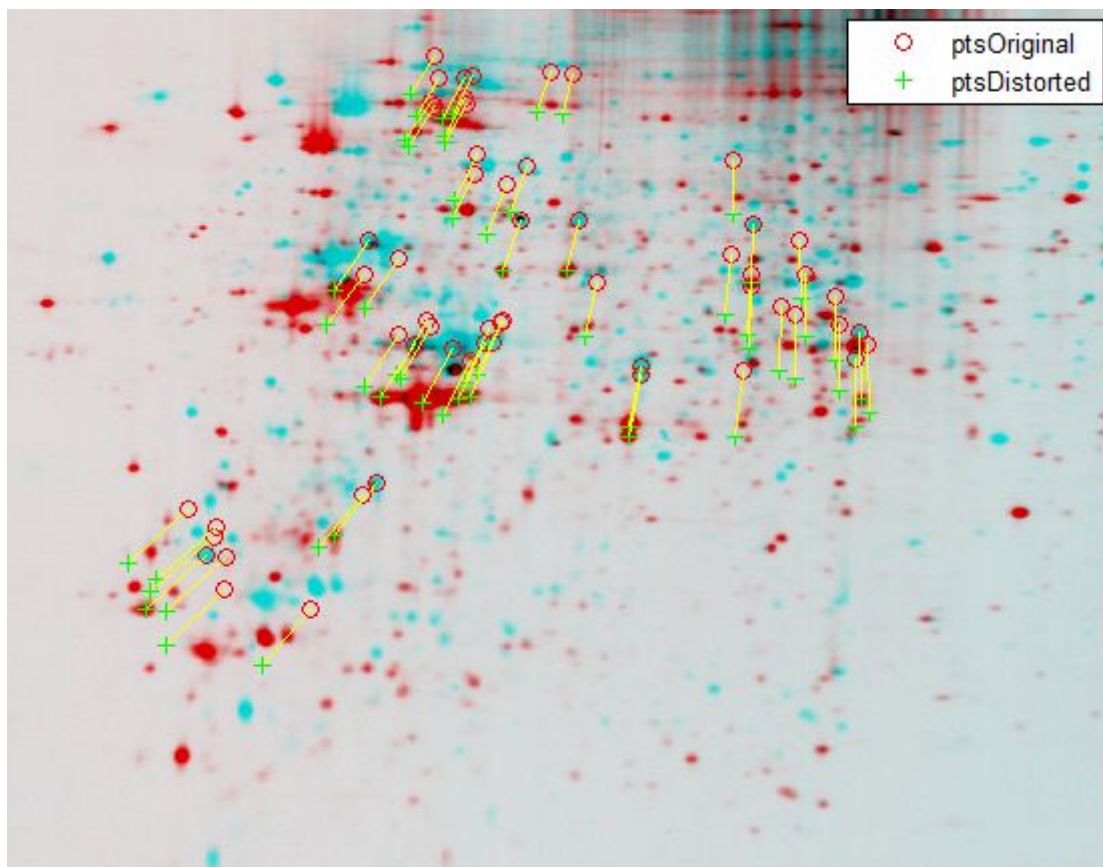


图 3.9 (b) 蛋白质点匹配结果 (去除异常点后)

图 3.9 基于 SURF 和 RANSAC 算法的蛋白质点匹配结果

在图 3.9 中, 用 “o” 表示参考 2-DE 图像中的蛋白质点, 用 “+” 表示待匹配 2-DE 图像中的蛋白质点, 对于匹配上的蛋白质点用直线进行连接。

表 3.1 给出了利用本章算法对 2-DE 图像蛋白质点匹配的结果。

表 3.1 本章算法匹配结果

检测到SURF特征点数目		有效SURF特征点数目	粗匹配点对数目 (含异常)	匹配点对数目 (去除异常)
参考图像	222个	222个	115对	57对
待匹配图像	201个	201个		

由第二章中表 2.1 可知, 参考 2-DE 图像中含有 393 个蛋白质点, 待匹配 2-DE 图像中含有 392 个蛋白质点。从表 3.1 的实验结果可以看出, 本章算法中检测到的点分别只有 222 和 201 个, 而且真正匹配上的蛋白质点只有 57 对。需要指出的是, SURF 算法在求主方向时过多依赖区域像素的梯度方向, 而这可能会导致所得的特征点主方向不准确; 而且后续特征向量的提取和匹配都依靠特征点主方向, 即便偏差角度不大也可能造成累积误差, 再加上后面 RANSAC 算法的迭代估计运算, 造成匹配不成功。

造成这个结果的另一个原因是，本章虽然采用的是 SURF 和 RANSAC 算法进行蛋白质点匹配，但是这些函数的使用却是对 MATLAB7.13 工具箱函数的调用。这就造成一个问题，由于这些工具箱函数的阈值及金字塔层的取值是固定的，而这造成的后果就是得到的参数及模型可能不是最合理的。而在实验时要求根据匹配结果的合理性来改变参数的值，直至出现一个合理的变换模型，完成蛋白质点对的匹配。

3.4 本章小结

为了解决第二章叙述的算法人工过多参与这一问题，在本章提出了一种基于 SURF 和 RANSAC 蛋白质点匹配算法。从实验的过程来看，人工过多参与这一问题得以解决；然而从实验的结果来看，由于本算法中主要是通过调用 MATLAB 工具箱函数来实现，所以结果的可信度比较低，而且只实现了少部分的蛋白质点匹配。造成这一结果的原因已经在 3.3 节中作了说明。根据本章算法的匹配结果可以发现本章算法具有以下优点：

- (1) SURF 具有计算速度快的特点，而且提高了特征检测和匹配的实时性；
- (2) RANSAC 能鲁棒的估计模型参数，它能从含有大量异常点的数据集中估计出精确的参数。

同时本章算法的缺点也是很明显的：

- (1) SURF 算法由于过度依赖区域像素梯度，这会造成得到的特征点主方向不准确，而且后续对特征向量的提取和匹配都要依据主方向；所以即使是不大的角度偏差，也会造成后续匹配的累积误差，造成匹配不成功；此外，图像金字塔的层取值也会影响尺度，造成尺度误差。

- (2) 如果对 RANSAC 算法的迭代次数设置上限，就会造成迭代的结果失真。RANSAC 算法得到可信结果的概率正比于迭代次数；此外，阈值的设置依赖于问题的具体要求。

第四章 不同匹配算法的对比

4.1 基于几何相似性的蛋白质点匹配算法

在本文的第二章中叙述了一种基于 Landmark 的蛋白质点匹配算法，在第三章中叙述了一种基于 SURF 和 RANSAC 的蛋白质点匹配算法。从前文的叙述可以知道，第二章的算法虽然匹配效果较好，但是需要大量的人工介入，当进行大量 2-DE 图像匹配时是不适用的；为了解决这一问题，在第三章叙述了一种自动匹配算法，然而从匹配的结果可以看出，第三章叙述的算法缺陷比较多，造成这一结果的原因主要是对算法理解的不够透彻。在实验中也做了其他方法的尝试，现在此处做一些简要的说明，希望对以后的工作能有一定的帮助。由于做的工作不够，所以下面叙述的 2-DE 图像匹配算法并不是很完整。

1、主要思路：

算法流程如图 4.1 所示。从分水岭算法检测蛋白质点的实验结果中提取出相关的数据，这些数据主要包含蛋白质点的面积、平均像素灰度、背景灰度以及蛋白质点的中心坐标值等，按照蛋白质点面积的大小，对这些蛋白质点数据进行重新排序（可以按由大到小的顺序排列），这样就可以得到两 2-DE 图像中蛋白质点的相关数据矩阵。然后，从参考图像的数据矩阵中提取出第一行（即面积最大），然后去和待匹配图像的数据矩阵进行面积和平均像素灰度的相似性计算，找出最相似的一个点，并在图上显示出这两个点，用同样的方法找出第二、第三个蛋白质点。通过计算这三个点构成的三角形的几何相似关系，如果这两个三角形是相同的（这里指的是角度和边长相似，可以认为是相同的），那么则确定这三个点是匹配的（如图 4.2 所示）。以此类推，通过改变参考图像中的一个点，去构造另外一个三角形，同样在待匹配图像中寻找相同的三角形。这样通过不断的变换三角形中的一个点，即可实现对一幅图像的遍历，通过不同三角形的匹配，从而实现所有蛋白质点的匹配。

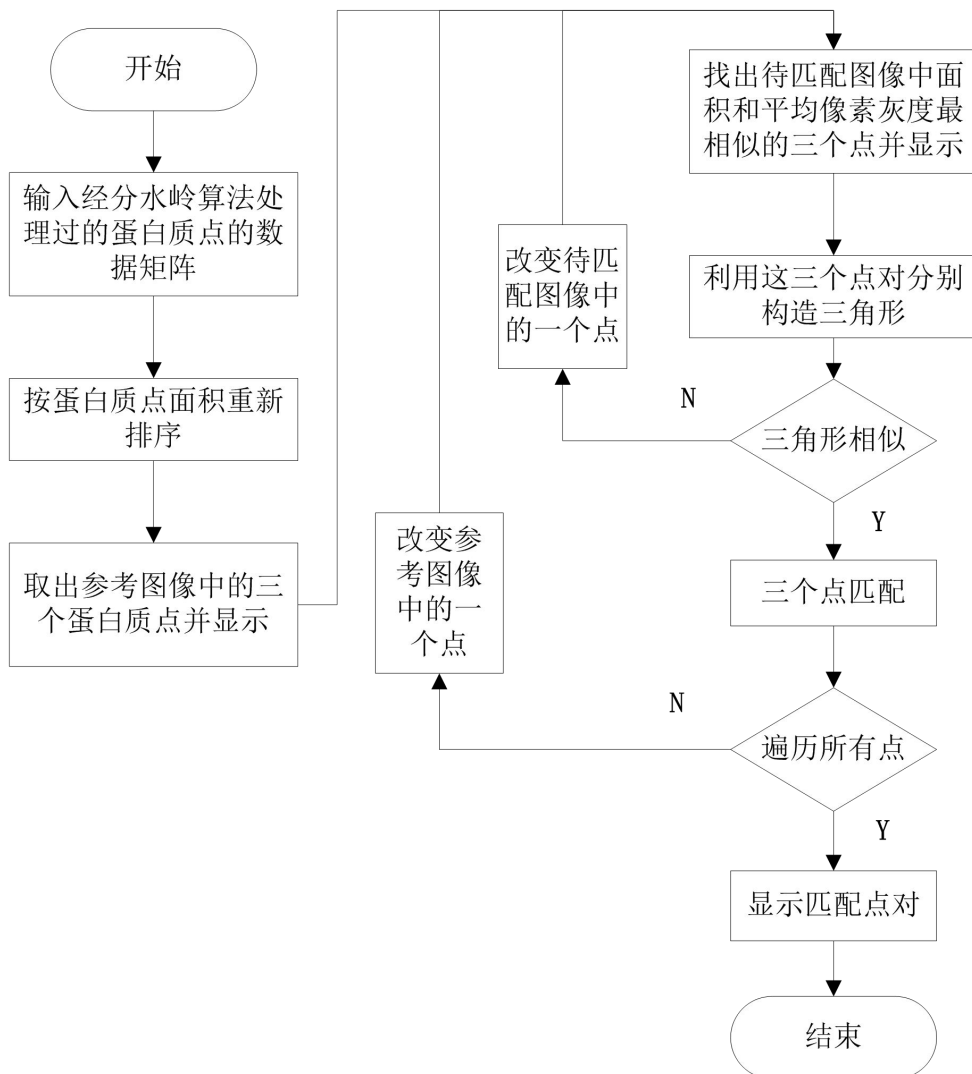
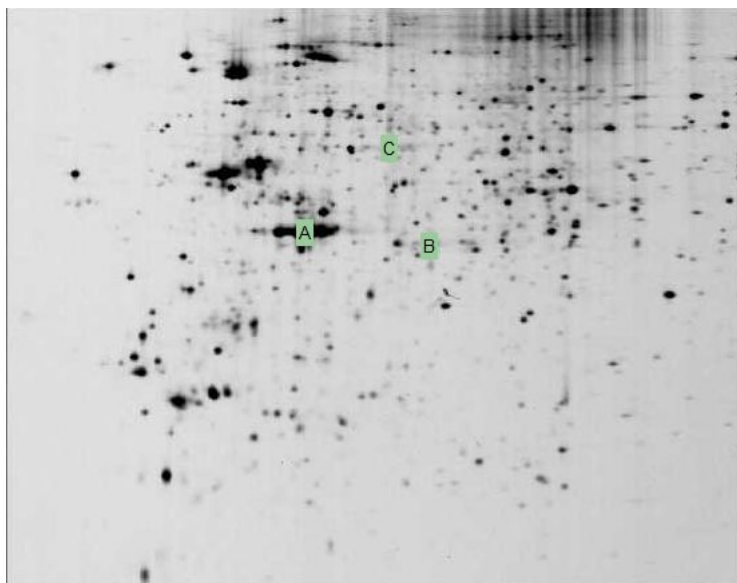
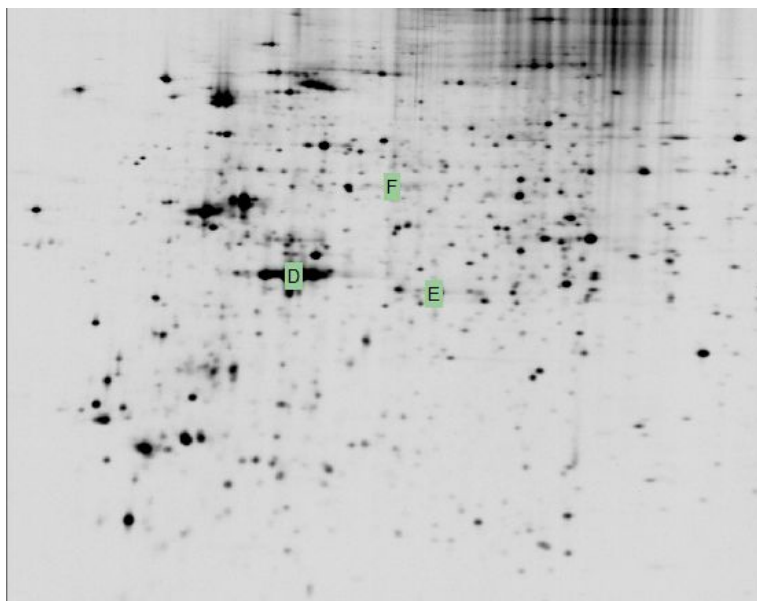


图 4.1 根据几何相似性实现蛋白质点匹配的算法流程图



(a) 参考 2-DE 图像中三个蛋白质点



(b) 待匹配 2-DE 图像中的三个蛋白质点

图 4.2 根据几何相似性实现匹配的三个蛋白质点

2、结果分析:

本算法可以实现对蛋白质点的匹配，但是可以发现，当 2-DE 图像中含有大量的蛋白质点时，算法的效率不是很高，计算量会很大。而且有一个比较大的缺点就是，在同一张 2-DE 图像中面积相同而且平均像素灰度相同的蛋白质点可能会有几个，这就在计算时会造成一定的误差，最终导致部分结果误匹配。

通过对这个实验中未能完成的匹配算法进行一个简要的叙述，希望能够对以后的工作提供一点不一样的思路。

4.2 三种不同算法的对比

在本文的第二章、第三章以及本章的 4.1 节分别提出了三种不同的 2-DE 图像蛋白质点匹配算法，从实验的结果可以看出来，虽然基于 Landmark 的蛋白质点匹配算法的人工介入过多，但是匹配的效果是这三种算法中最好的；第三章中叙述的自动匹配算法解决了第二章中人工介入过多的问题，然而由于利用的 SURF 函数是 MATLAB 工具箱封装好的函数，所以未能根据本实验中 2-DE 图像的实际情况设计合适的参数，而这使得后面的 RANSAC 算法进行迭代计算时产生一个非最合理的变换模型，最终导致实验结果不是最准确的，即蛋白质点匹配的成功率比较低；而在本章 4.1 节中提出的基于几何相似的匹配算法，虽然免去了人工介入，也能实现蛋白质点的匹配，但是可以发现，如果一幅 2-DE 图像中蛋白质点的面积和平均像素灰度是极度相似的，则该算法是无法实现蛋白质点的

匹配的，而且该算法的计算量相比于前面两章叙述的算法要大。表 4.1 给出了三种不同匹配算法的对比情况。

表 4.1 三种不同匹配算法的对比

	基于Landmark的蛋白点匹配算法	基于SURF和RANSAC的蛋白点匹配算法	基于几何相似性的蛋白点匹配算法
计算量	由于调用了工具箱函数cp2tform进行变换模型推断，所以计算量较小。	RANSAC进行迭代运算估计模型参数时调用的是工具箱函数，所以计算量较小。	计算量较大
自动化程度	半自动匹配	自动匹配	自动匹配
匹配结果	手动添加的Landmark标记点数越多，蛋白点匹配成功率越高。	由于调用的是工具箱函数，只匹配了少量的蛋白点，但是精度高。	尚未完成大量的蛋白点匹配

第五章 总结

本文以图像匹配特征、坐标变换模型以及统计学为基础，以 2-DE 图像蛋白质点为匹配研究对象，重点叙述了 2-DE 图像间蛋白质点匹配方法：（1）基于 Landmark 的蛋白质点匹配算法；（2）基于 SURF 和 RANSAC 的蛋白质点匹配算法；（3）基于几何相似性的匹配算法（此为 4.1 节中叙述的匹配算法）。

（1）通过在 2-DE 图像上添加 Landmark 标记点，结合传统的 ICP 算法，提出了一种基于 Landmark 的蛋白质点匹配算法。实验结果表明该算法简单有效，利用手动标记特征蛋白质点（尤其是对图像形变处添加标记点），并调用 MATLAB7.13 工具箱函数 `cp2tform` 推断几何变换参数，并确立 Landmark 标记点间的几何变换模型，从而利用这个模型对两 2-DE 图像建立统一的空间坐标系，进行 2-DE 图像间的蛋白质点匹配。通过比较不同数量的 Landmark 标记点来检验本算法的匹配效果，从匹配结果图可以发现，当 Landmark 标记数越多而且在凝胶图像上分布越均匀的时候，匹配效果越好。

（2）之所以提出基于 SURF 和 RANSAC 的蛋白质点匹配算法，主要是为了解决第二章算法中人工介入较多的问题。在该算法中首先通过调用 MATLAB7.13 工具箱函数 `detectSURFFeatures` 进行特征点寻找，然后利用 `extractFeatures` 函数进行特征点特征提取，最后用 `matchFeatures` 函数得到蛋白质点匹配点集（这里的匹配结果包含大量的异常点）；在 RANSAC 算法中通过迭代方式估计数学模型，并利用这个合理的模型去除异常点，得到一个只包含局内点的匹配点集。从实验结果可以看出，在本文叙述的算法中，匹配的结果不是很好。这主要是由调用工具箱函数造成的，调用工具箱函数时不能改变封装函数的参数及阈值，所以匹配出来的效果不是很好。

（3）在本文的 4.1 节提出了一种基于几何相似性的匹配算法，但是由于该算法的匹配实验并没有全部完成，所以只是提出了一些算法思路。从算法的思路可以看出，该算法可以实现蛋白质点的匹配；但是可以发现算法的计算量很大，而且对蛋白质点分水岭检测的结果要求很高。当 2-DE 图像中含有大量蛋白质点且蛋白质点面积及平均像素灰度相似时，该算法的计算量是非常巨大的。

经过这次毕设的学习，使自己对数字图像处理原理有了一个基本的认知。通过对数字图像处理的学习，了解到了当前的一些图像匹配技术，这些图像匹配技术对本文中叙述的 2-DE 图像蛋白质点算法起到了至关重要的作用。然而通过自己对本文中叙述的匹配算法进行编程实现时发现，要彻底掌握一种匹配算法，需

要对整个数字图像处理技术有一个深入的理解，而这需要长期的学习和实践。所以说，只有学习了大量的图像匹配算法之后，才能够根据各个算法的优缺点提出相应的改进算法。通过对这些算法的学习以及编程实现，可以发现并不是所有的匹配算法都适合 2-DE 图像蛋白质点匹配，所以在匹配不同类型的图像时，应按要求采用合适的算法去完成匹配。本章所叙述的这些 2-DE 图像蛋白质点匹配算法，是对现有算法的一个演示过程，希望能够对以后的算法学习起到一定的参考作用。

参考文献

- [1]. Marc Wilkins. PROTEOME is the PROTEINS expressed by a genome or a tissue[J]. Electrophoresis, 1995, 07.
- [2]. 冯雪, 王彬. 蛋白质组学研究的相关技术进展[J]. 生物学教学, 2010, (3): 4-5.
- [3]. 张曼. 蛋白质组学研究中的核心技术 - 双向凝胶电泳[J]. 现代医学, 2004, 32(4): 211-213.
- [4]. 宋革, 姜勇. 二维凝胶电泳的新技术及其应用[J]. 中国微循环, 2005, 9(1).
- [5]. Kameyama K, Kishi Y, Yoshimura M, et al. Tyrosine phosphorylation in plant bending[J]. Nature. 2000, 407: 37.
- [6]. 熊邦书. 基于分层策略的凝胶图像间蛋白点匹配算法[J]. 中国生物医学工程报, 2012, 31(3): 422-427.
- [7]. Khan N, Rahman S. A new approach to detect similar proteins from 2D gel electrophoresis image[C]. Proceedings of 3rd IEEE Symposium on Bioinformatics and Bioengineering, 2003: 182-189.
- [8]. Rogers M, Graham J, Tonge R P. 2-D electrophoresis gel registration using feature matching[J]. IEEE International Symposium on Biomedical Imaging, 2004, 2(8): 1436-39.
- [9]. Toru Wakahara, Kazumi Odaka. Adaptive Normalization of Handwritten Characters Using Global/Local Affine Transformation[J]. IEEE Trans, 1998, 20(12): 1332-1341.
- [10]. Grauman K, Darrell T. The pyramid match kernel: discriminative classification with set of image features[C]. Proceedings of the 10th IEEE International Conference on Computer Vision, 2005: 1458-1465.
- [11]. 罗钟铨, 刘成明. 灰度图像匹配的快速方法[J]. 大连理工大学学报, 2005.
- [12]. Minh-Tuan T H, Hieu T H, Nguyen H. Two-Step Iterative Registration for 2D-Gel Electrophoresis Images[J]. IEEE, 2007, 7(4): 274-280.
- [13]. Rafael C Gonzalez, Richard E Woods, Steven L Eddins(著), 阮秋琦(译). 数字图像处理的 MATLAB 实现(第 2 版)[M]. 清华大学出版社, 2013.
- [14]. David G Lowe. Distinctive Image Features from Scale-Invariant Key points[J]. Computer Vision, 2004.
- [15]. Fischler M A, Bolles R C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Communications of the ACM, 1981, 24(6): 381-395.

致谢

经过近个三个月的努力，终于完成了本次毕设的实验和论文部分。首先要衷心感谢我的指导老师欧巧凤老师，在本次毕设过程中，欧老师给予我悉心的指导，为我答疑解惑，并提供了许多宝贵的建议。在毕设的初期准备阶段，欧老师不辞辛苦的为我解释课题的内容及意义，并提供了许多宝贵的参考书籍，这使得我能够在较短的时间内了解到我的毕设主要工作方向。在算法编程实现阶段，欧老师给予了大量的设计思路，并对我提出的想法给予细致的分析，使我能够更加透彻的理解我的课题，从而能够更好的完成我的毕设。在修改论文期间，欧老师不辞辛苦的一次次为我指出论文中的错误之处，并提供相应的解决方法。就是欧老师这种严谨的治学态度才使得我能够顺利的完成本次毕设，再次向欧老师致以崇高的敬意。

此外，还要感谢施歌、刘文苗等同学，感谢你们在我的算法编程实现过程中给予的宝贵建议，祝你们学业有成。

最后，再次感谢所有在本次毕设过程中对我给予帮助和关心的朋友们，祝大家身体健康，事事顺心！

附录

```

%找到stock图像中的SURF特征点
ptsOriginal = detectSURFFeatures(fdat1);
ptsDistorted = detectSURFFeatures(fdat2);
%特征提取
[featuresIn validPtsIn] = extractFeatures(fdat1, ptsOriginal);
[featuresOut validPtsOut] = extractFeatures(fdat2, ptsDistorted);
%特征匹配
index_pairs = matchFeatures(featuresIn, featuresOut);
%检索每个图像的对应点的位置
matchedOriginal = validPtsIn(index_pairs(:,1));
matchedDistorted = validPtsOut(index_pairs(:,2));
%显示推定的匹配点对，这里包括外点（即异常点）
cvexShowMatches(fdat1,fdat2,matchedOriginal,matchedDistorted);
title('Putatively matched points (including outliers)');
%从一组包含“局外点”的数据集中，通过迭代方式估计数学模型。
%运用统计学上的随机抽样一致性（RANSAC），找到合适的变换矩阵来匹配点
%对，并去除异常点。
gte = vision.GeometricTransformEstimator; %几何估计变换，默认为 RANSAC
gte.Transform = 'projective';
gte.NumRandomSamplingsMethod = 'Desired confidence';%随机抽样方法 Desired
%confidence
gte.MaximumRandomSamples = 1000;%随机样本数
gte.DesiredConfidence = 99;
[tform_matrix
inlierIdx]=step(gte,matchedDistorted.Location, ...matchedOriginal.Location);
%inlierIdx=57局内点
cvexShowMatches(fdat1,fdat2,matchedOriginal(inlierIdx),...
matchedDistorted(inlierIdx),'ptsOriginal','ptsDistorted');
title('Matching points (inliers only)');

```