

像对匹配 方法研究

贾迪 朱红 著



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

国家自然科学基金资助项目 (61601213)

像对匹配方法研究

贾迪朱红著



电子工业出版社

Publishing House of Electronics Industry

北京 • BEIJING

内 容 简 介

本书对近年来的像对匹配方法予以全面总结,并针对宽基线条件下像对匹配问题给出一些解决方法,这些方法均为作者近年来的研究成果。本书共4章,第1章介绍近年来的像对匹配方法,主要内容包括像对局部不变特征点匹配方法、像对直线特征匹配方法、像对区域特征匹配方法,并给出这些匹配方法的综合对比与分析;第2章介绍局部不变特征点稠密匹配方法,主要内容包括基于Deep Matching的像对高效稠密匹配方法、平滑约束与三角网等比例剖分像对稠密匹配方法;第3章为直线特征匹配与提纯,主要内容包括直线特征匹配、面向图像的直线特征矫正与匹配结果提纯;第4章是模板特征选取与匹配,主要内容包括像对模板选择与匹配、提高模板匹配性能的方法。

本书可作为图像匹配、配准研究方向硕士生、博士生、专业教师的科研参考用书,也可为广大计算机视觉处理的科技工作者提供技术参考。

未经许可,不得以任何方式复制或抄袭本书之部分或全部内容。
版权所有,侵权必究。

图书在版编目(CIP)数据

像对匹配方法研究 / 贾迪, 朱红著. —北京: 电子工业出版社, 2019.7
ISBN 978-7-121-36686-4

I. ①像… II. ①贾… ②朱… III. ①图象处理—研究 IV. ①TP391.413

中国版本图书馆 CIP 数据核字 (2019) 第 103684 号

策划编辑: 刘小琳

责任编辑: 刘小琳 特约编辑: 刘 炯 等

印 刷: 北京七彩京通数码快印有限公司

装 订: 北京七彩京通数码快印有限公司

出版发行: 电子工业出版社

北京市海淀区万寿路 173 信箱 邮编: 100036

开 本: 720×1000 1/16 印张: 9.75 字数: 145 千字

版 次: 2019 年 7 月第 1 版

印 次: 2019 年 7 月第 1 次印刷

定 价: 68.00 元

凡所购买电子工业出版社图书有缺损问题, 请向购买书店调换。若书店售缺, 请与本社发行部联系, 联系及邮购电话: (010) 88254888, 88258888。

质量投诉请发邮件至 zlts@phei.com.cn, 盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式: liuxl@phei.com.cn, (010) 88254538。

作者简介

贾迪，男，博士，辽宁工程技术大学副教授，中国图像图形学会会员，亚利桑那州立大学访问学者。主要研究方向为图像匹配与三维重建，计算机视觉定位。

朱红，女，博士，防灾科技学院讲师，自然资源部国土卫星遥感应用中心，主要从事图像匹配，遥感影像超分辨率重建以及卫星姿态探测的相关研究工作。

前言



像对匹配是指能够自动识别存在于不同像平面上的特征基元所对应的同物方结构信息。该技术作为计算机视觉的核心任务，是后续高级图像处理的关键，如图像拼接、三维重建、影像融合、超分辨率重建、视觉定位等，所涉及的领域有工业检测、导弹的地形匹配、光学和雷达的图像跟踪、交通管理、工业流水线的自动监控、工业仪表的自动监控、医疗诊断、资源分析、气象预报、文字识别及图像检索等。从摄影测量角度可将立体像对分为窄基线与宽基线两种获取模式，目前窄基线像对的匹配方法已较为成熟，而在宽基线条件下由于获取同一目标影像的设备、时间、视点、视角、光照条件等因素均不相同，故如何提高这类像对匹配的执行速度、配准率及鲁棒性成为目前的研究热点。

本书中给出的方法为作者近年来的研究成果，主要研究内容为在宽基线条件下的像对匹配方法。在局部不变特征点稠密匹配的研究中，基于 HOG 特征的稠密化估计给出一种稀疏到稠密宽基线像对匹配方法，适用于立体像对匹配；提出密度聚类平滑约束提纯内点及三角网等比例剖分稠密匹配，避免由于某些局部外点造成仿射变换矩阵估计不准确而影响整体平面稠密匹配准确率的问题。在直线特征匹配与提纯方法的研究中，给出三种直线特征匹配方法，包括重合度约束直线特征匹配、线段元支撑区主成分相似性约束特征线匹配、多重约束下的直线特征匹配；给出建立梯度引力图的方法，并以此为基础对直线特征位置进行矫正，结合极限约束提纯直线匹

配结果。在区域匹配方法的研究中,给出结合 SV-NCC 度量彩色图像块间相似性的方法,并通过分值图选择最佳模板匹配位置;给出两种提高模板匹配性能的方法,分别为缩小模板匹配的搜索空间与无纹理区域的纹理构造方法。

本书共 4 章,第 1 章对近年来的像对匹配方法予以总结,主要内容包括像对局部不变特征点匹配方法、像对直线特征匹配方法、像对区域特征匹配方法,并给出这些匹配方法的综合对比与分析;第 2 章介绍局部不变特征点稠密匹配方法,主要内容包括基于 Deep Matching 的像对高效稠密匹配方法、平滑约束与三角网等比例剖分像对稠密匹配方法;第 3 章为直线特征匹配与提纯,主要内容包括直线特征匹配、面向图像的直线特征矫正与匹配结果提纯;第 4 章是模板特征选取与匹配,主要内容包括像对匹配的模板选择与匹配、提高模板匹配性能的方法。

参与本书相关科学实验与内容校对的同学包括朱宁丹、赵明远、吴思、李玉秀、杨宁华,在此对他们表示感谢!

由于个人水平所限,书中难免会出现疏漏和失误,恳请广大读者批评指正,并提出宝贵意见。

著 者

目 录



第 1 章 像对匹配方法综述.....	1
1.1 像对局部不变特征点匹配方法.....	2
1.2 像对直线特征匹配方法.....	20
1.3 像对区域特征匹配方法.....	23
1.4 像对匹配方法综合对比分析.....	33
参考文献	40
第 2 章 局部不变特征点稠密匹配.....	49
2.1 基于 Deep Matching 的像对高效稠密匹配方法	49
2.1.1 基于 HOG 特征的稠密化估计.....	51
2.1.2 像对稠密匹配方法	52
2.1.3 实验结果与分析	56
2.2 平滑约束与三角网等比例剖分像对稠密匹配方法	63
2.2.1 密度聚类平滑约束提纯内点	67
2.2.2 三角网等比例剖分稠密匹配	69
2.2.3 实验结果与分析	73
2.3 本章小结	79
参考文献	79

第 3 章 直线特征匹配与提纯	84
3.1 直线特征匹配	84
3.1.1 重合度约束直线特征匹配方法	85
3.1.2 线段元支撑区主成分相似性约束特征线匹配方法	92
3.1.3 多重约束下的直线特征匹配方法	100
3.2 直线特征矫正与匹配结果提纯	108
3.2.1 梯度引力图	109
3.2.2 直线位置矫正方法	111
3.2.3 提纯直线匹配结果	113
3.2.4 实验结果与分析	117
3.3 本章小结	123
参考文献	123
第 4 章 模板特征选取与匹配	127
4.1 像对模板选择与匹配	127
4.1.1 SV-NCC 度量彩色图像块间的相似性	128
4.1.2 分值图与选择最佳模板匹配位置	131
4.1.3 实验结果与分析	133
4.2 提高模板匹配性能的方法	138
4.2.1 缩小模板匹配的搜索空间	138
4.2.2 无纹理区域的纹理构造方法	141
4.3 本章小结	147
参考文献	147

第 1 章

01

像对匹配方法综述

图像匹配作为计算机视觉的核心任务，是后续高级图像处理的关键，如目标识别、图像拼接、三维重建、视觉定位、场景深度计算等。近年来，尽管国内外学者在该领域的研究中提出许多优秀的方法，但还未见对这类方法予以全面总结的文章，为此，本章从局部不变特征点、直线、区域匹配三个方面对这些方法予以综述。局部不变特征点匹配在图像匹配领域出现最早，本章对这类方法中的经典方法仅予以简述，对于近年来新出现的方法予以重点介绍，尤其是基于深度学习的匹配方法，包括 TILDE、Quad-networks、Deep Desc、LIFT 等方法。由于外点剔除类方法常用于提高局部不变点特征匹配的准确率，因此也对这类方法予以介绍，包括 BF、GMS、VFC 等。与局部不变特征点相比，线包含更多场景和对象的结构信息，更适用于具有重复纹理信息的像对匹配。线匹配的研究需要克服包括端点位置不准确、线段外观不明显、线段碎片等问题，解决这类问题的方法有 LBD、CA、LP、共面线点投影不变量法等，本章从问题解决的角度对这类方法予以介绍。区域匹配从区域特征提取与匹配、模板匹配两个角度对这类方法予以介绍。典型的区域特征提取与匹配方法包括 MSER、TBMR，模板匹配方法包括 FAST-Match、CFast-Match、DDIS、OATM，以及深度学习类的方法如 MatchNet、L2-Net、PN-Net、Deep CD 等。本章从局部不变特征点、直线、区域三个方面对图像匹配方法予以总结对比，包括不同

匹配方法计算时间与准确率的比较、基于深度学习类匹配方法的比较等，给出这类方法对应的论文及代码下载地址，并对未来的研究方向予以展望。图像匹配是计算机视觉领域后续高级处理的基础，目前在宽基线匹配、实时匹配方面仍需要进一步研究。

早期图像匹配方法以角点检测与匹配为主，从 Harris 角点检测算子^[1]到 FAST 检测算子^[2]，以及对这类角点检测算子的改进方法。SIFT^[3]方法的提出将研究者的思维从角点检测中解放出来，是迄今为止该方向引用量最多的技术。目前，SIFT 类方法不再占据主导地位，基于深度学习的图像匹配方法逐步兴起，这类图像匹配不再依据研究者的观察和专业知识，而依靠数据的训练，因此匹配精确度更高。文献[4]对 SIFT 及其改进方法予以总结，文献[5]对局部图像描述符进行了综述，然而在图像匹配的诸多方法中，局部不变特征点匹配仅是其中的一类方法，直线匹配与区域匹配这两类方法还未见文章予以总结，为此本章从补充图像匹配方法的角度出发，将图像匹配方法分为三大类予以总结：局部不变特征点匹配、直线匹配、区域匹配，目的是为该方向的研究学者提供更为全面的文献综述，同时也为即将开展该方向的研究人员提供参考。

1.1 像对局部不变特征点匹配方法

局部不变特征点匹配在图像匹配领域发展最早，一幅图像的特征点由两部分组成：关键点和描述子。关键点是指特征点在图像中的位置，具有方向、尺度等信息；描述子通常是一个矢量，用于描述关键点邻域的像素信息。在进行局部不变特征点匹配时，通常只需要在矢量空间对两个描述子进行比较，距离相近则判定为同一个特征点，角点、边缘点等都可以作为潜在特征点。在角点检测方法中最常用的方法是基于图像灰度的方法，

如 Harris 方法^[1]。Harris 方法通过两个正交方向上强度的变化率对角点进行定义,其存在尺度固定、像素定位精度低、伪角点较多、计算量大等问题。为此,诸多学者提出相应改进方法^[6-8]。文献[6]将多分辨率思想引入 Harris 角点,解决了 Harris 方法不具有尺度变化的问题。文献[7]在 Harris 方法中两次筛选候选点集,利用最小二乘加权距离法实现角点亚像素定位,大幅度提高角点检测效率和精度。文献[8]将灰度差分及模板与 Harris 方法相结合,解决了 Harris 方法存在较多伪角点、计算量大等问题。FAST 方法^[2]通过邻域像素对比进行特征点检测,并引入机器学习加速过程,可应用在对实时性要求较高的场合,如视频监控中的目标识别。由于 FAST 方法仅处理单一尺度图像,并且检测的不仅是“角点”这一特征,还可以检测其他符合要求的特征点,如孤立的噪点等。当图像中噪点较多时会产生较多外点,导致图像鲁棒性下降。

SIFT 方法^[3]的提出打破了角点检测的僵局,使特征点检测不再局限于角点检测,后续相继提出针对 SIFT 方法的改进方法。Xu 等人^[5]对局部图像描述符进行分析描述,对这类方法的计算复杂度、评价方法和应用领域予以总结。刘立等人^[4]对 SIFT 方法的演变及在不同领域的典型应用进行了较为全面的论述,并比较了各类方法的优缺点。随后,针对 SIFT 方法时间复杂度高的问题相继提出了 PCA-SIFT^[9]、SURF^[10]、SSIF^[11],以及对彩色图像进行处理的 CSIFT^[12]、使用对数极坐标分级结构的 GLOH^[13]、具有仿射不变性的 ASIFT^[14]等。

以上特征点匹配方法均基于人工设计的特征点检测器,深度学习的快速发展使其在图像匹配领域的应用成为现实。在局部特征点的重复检测方面,FAST-ER 方法^[15]把特征点检测器定义为一种检测高重复点的三元决策树,并采用模拟退火方法对决策树进行优化,从而提高检测重复率。由于在每次迭代过程中,都需要对重新应用的新决策树进行检测,并且其性能受到初始关键点检测器的限制,降低了该方法的鲁棒性。Verdie 等人^[16]提出 TILDE 像对匹配方法,该方法能够较好地对由天气、季节、时间等因素

引起的剧烈光照变化情况下的可重复关键点进行检测。参与训练的候选特征点由多幅训练图像中采用 SIFT 方法提取的可重复关键点组成,如图 1.1 (a) 所示;正样本是以这些点为中心的区域,负样本是远离这些点的区域。在进行回归训练时,正样本在特征点位置返回最大值,在远离特征点位置则返回较小值,如图 1.1 (b) 所示;在回归测试时,将测试图像分成固定大小的图像块,其回归响应如图 1.1 (c) 所示,再根据非极大值抑制提取特征点,如图 1.1 (d) 所示。该方法适用于处理训练数据和测试数据为同一场景的图像。

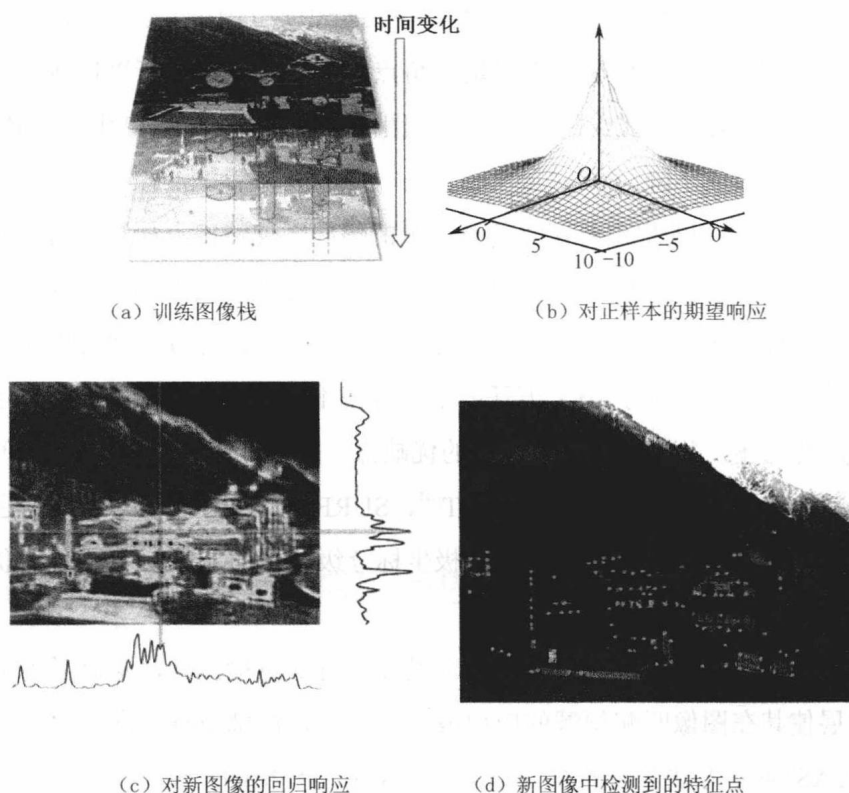


图 1.1 TILDE 像对匹配方法概述

一个良好的局部特征检测器应具备两个特性:①检测可区分的特征;②协变约束,即在不同的变换下重复检测一致特征。而大多数检测器都只

考虑其中一个特性，如 TILDE 像对匹配方法采用手动标记的数据作为区分性特征训练。Zhang 等人^[17]综合考虑这两个特性，提出基于学习的协变特征检测器。该方法将 TILDE 像对匹配方法的输出作为候选标准图像块，通过变换预测器的训练建立学习框架，将局部特征检测器的协变约束转化为变换预测器的协变约束，以便利用回归（如深度神经网络）进行变换预测。预测的变换有两个重要性质：①变换的逆矩阵能将观察到的图像块映射到“标准块”，“标准块”定义了具有可区分性的图像块及块内“典型特征”（如单位圆）的位置和形状；②将变换应用到“典型特征”可以预测图像块内变换特征的位置和形状。变换预测网络的训练流程如图 1.2 所示，变换矩阵 g_i 应用于标准图像块 $\bar{x}_i$ ，得到变换图像块 $g_i \times \bar{x}_i$ ；两个图像块经过变换预测器，分别输出回归量 $\phi(\bar{x}_i)$ 和 $\phi(g_i \times \bar{x}_i)$ ；将变换 g_i 应用于 $\phi(\bar{x}_i)$ 得到 $g_i \times \phi(\bar{x}_i)$ ，并利用矩阵 Frobenius 范数计算协变约束损失。网络参数设置如下：第一层卷积核大小为 5×5 ，具有 32 个输出通道，其后是 2×2 的最大池化层；第二层卷积核大小为 5×5 ，具有 128 个输出通道，其后是 2×2 最大池化层；第三层卷积核大小为 3×3 ，具有 128 个输出通道；第四层卷积核大小为 3×3 ，具有 256 个输出通道；最后一层卷积核大小为 1×1 ，其输出通道数量和回归变换的参数数量相同。此外，所有卷积层均使用 ReLU 作为激活函数。该检测器可以应用在图像搜索和场景重建中，但需要一组可靠关键点作为训练输入。

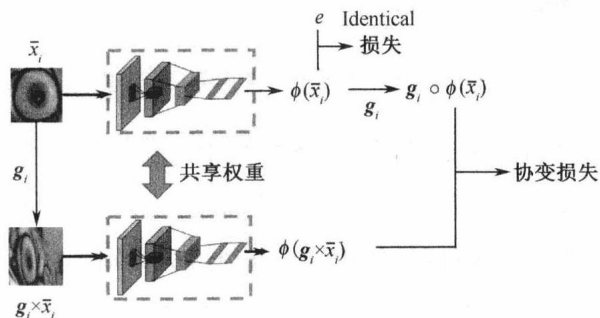


图 1.2 变换预测网络的训练流程

基于监督学习的图像匹配方法以人工设计为基础,如 TILDE 相对匹配方法使用 DOG 收集训练集,这些方法对于跨模态任务(如 RGB/深度模态对)可能不再适用。

Savinov 等人^[18]提出 Quad-networks,采用无监督学习方式进行特征点检测。该方法将关键点检测问题转化为图像变换上的关键点一致性排序问题,优化后的排序在不同的变换下具有重复性,其中关键点来自响应函数的顶/底部分位数。Quad-networks 的训练过程如图 1.3 所示,在两幅图像中提取随机旋转像块对 (1, 3) 和 (2, 4); 每个像块经过神经网络输出一个实值响应 $H(p|w)$, 其中, p 表示点, w 表示参数矢量; 通过四元组的排序一致函数计算铰链损失,并通过梯度下降法优化。该网络对 RGB 图像和深度图像采用不同的层参数,采用元组对层参数表示: $c(f, i, o, p)$ 表示卷积层,其卷积核大小为 $f \times f$, i 为输入通道数, o 为输出通道数, p 表示图像边界的 p 个像素采用零填充。Quad-networks 在 RGB/RGB 模式和 RGB/深度模式的重复检测性能均优于 DOG,可以和基于学习的描述符相结合进行图像匹配,还可用于视频中的兴趣帧检测。

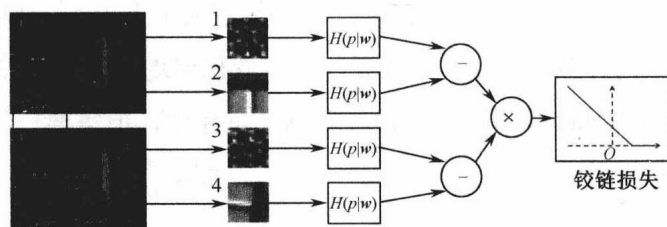


图 1.3 Quad-networks 训练过程

基于深度学习的方法不仅可以学习特征检测器,还可用于对特征描述符进行学习。Simo-Serra 等人^[19]提出用于特征点描述符判别学习的 Deep Desc。该方法采用 Siamese 网络侧重训练难以区分类别的样本,输入图像块对,将 CNN 输出的非线性映射作为描述符,采用欧氏距离计算相似性并最小化其铰链损失。该方法适用于不同的数据集和应用,包括宽基线图像匹配、非刚性变形和极端光照变化的情况,但该方法需要大量的训练数据来

保证其鲁棒性。

以上基于学习的匹配方法大部分对匹配过程中的某个阶段单独进行操作。Yi 等人^[20]提出的 LIFT 结合空间变换网络^[21]和 Softargmax 函数，将基于深度学习的特征点检测^[16]、基于深度学习的方向估计^[22]和基于深度学习的描述符^[19]连接成一个统一网络，从而实现完整特征点匹配处理流水线，如图 1.4 所示。其中，图像块的裁剪和旋转通过空间变换网络实现，训练阶段采用四分支 Siamese 体系结构（见图 1.5），输入特征点所在图像块，其位置和方向均来自 SFM 方法的输出，其中 P^1 和 P^2 来自同一个 3D 点在不同视角下的图像， P^3 为不同 3D 点投影的图像块， P^4 为不包含任何特征点的图像块， S 代表得分图， X 代表特征点位置。

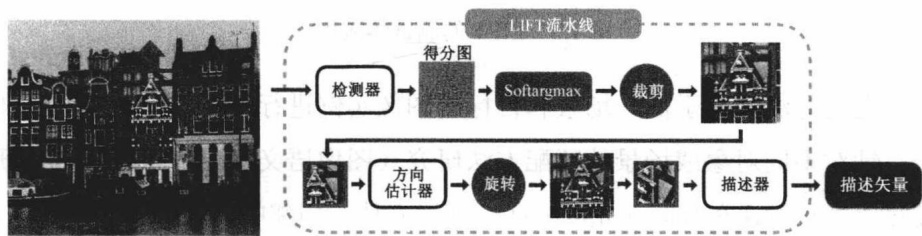


图 1.4 LIFT 集成特征提取流水线

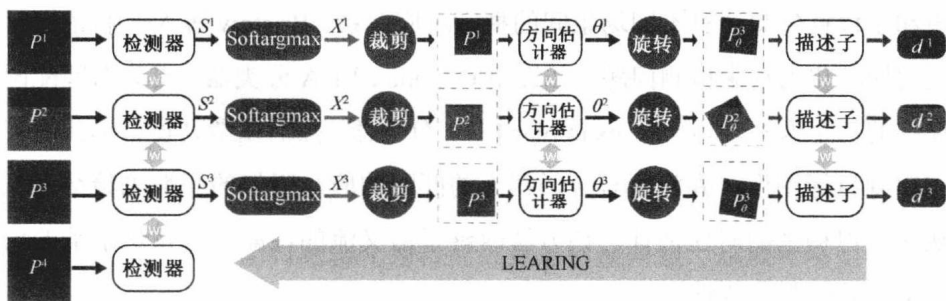


图 1.5 四分支 Siamese 体系结构

采用从后至前的训练策略，即先训练描述子，再训练方向估计，最后训练特征点检测。在测试阶段，将特征点检测与方向估计及描述子分开，

使优化问题易于处理。如图 1.6 所示，输入多尺度图像，以滑窗形式进行特征点检测，提取局部块逐个分配方向，再计算描述子。与 SIFT 相比，LIFT 能够提取出更为稠密的特征点，并且对光照和季节变化具有很高的鲁棒性。

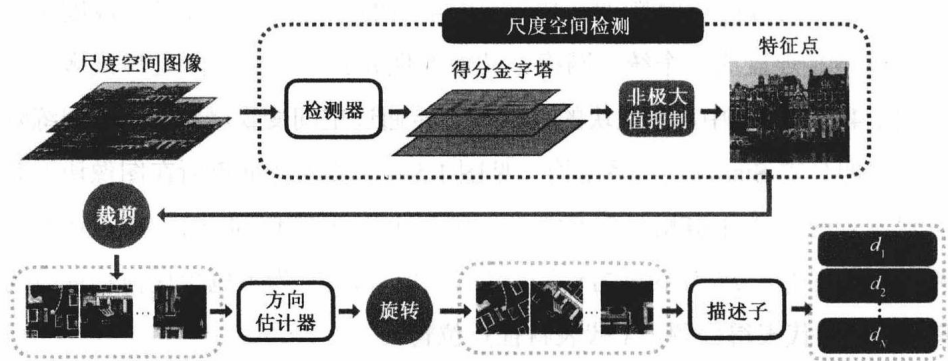


图 1.6 LIFT 的测试体流程

上述方法均基于同一场景和目标的图像实例进行匹配方法研究，近年来，针对不同对象或场景的匹配方法研究（图像语义匹配）正逐步成为研究热点。与考虑在时间（光流）或空间（立体）相邻的图像特征对应不同，语义对应的特征是图像具有相似的高层结构，而其精确的外观和几何形状可能不同。经典 SIFT 流方法^[23]提出不同场景的稠密对应概念，通过平滑约束和小位移先验计算不同场景间的稠密对应关系。Bristow 等人^[24]将语义对应问题转化为约束检测问题，并提出 exemplar-LDA 分类器。首先对匹配图像中的每个像素学习一个 exemplar-LDA 分类器，然后以滑动窗口形式将其应用到目标图像，并将所有分类器上的匹配响应与附加的平滑先验结合，从而获得稠密的对应估计。该方法改善了语义流的性能，在背景杂乱的场景下具有较强鲁棒性。

上述两种方法将流概念推广到仅在语义上相关的图像对，都对属于相同对象类的像对进行匹配，而对属于不同对象类别的图像则不再适用。Novotny 等人^[25]提出基于几何敏感特征的弱监督学习方法 AnchorNet。在只有图像级标签的监督下，AnchorNet 依赖一组从残差超列 HC（Hypercolumn）

中提取具有正交响应的多样过滤器，该过滤器在同一类别的不同实例或两个相似类别之间具有几何一致性。AnchorNet 通过在 ILSVRC12 上预先训练的 ResNet 50 模型初始化网络参数，并采用两阶段优化与加速训练完成匹配，如图 1.7 所示。第一阶段学习具有可区分性和多样性的类特定特征。输入图像经过 ResNet 网络后，对 Res2c、Res4c 和 Res5c 层的矫正输出进行升采样和级联，到 $56 \times 56 \times 768$ 的 HC 张量，级联之前的 PCA 用于压缩每层提取的描述符至 256 维， L_2 归一化用于能量平衡；可区分性损失 L_{Discr} 通过全局最大池标识每个过滤器的最强响应， $L_{\text{Discr}}^{\text{aux}}$ 通过全局平均池抑制负样本图像的响应；多样性损失 L_{Div}^A 和 L_{Div}^B 可以增强滤波器的正交响应。第二阶段学习类无关（Class-agnostic Features）特征，采用自动编码器对第一阶段获得的特征进行压缩，使得在类间可以共享这些特征，并最小化重建损失。AnchorNet 提高了跨类语义匹配的性能。

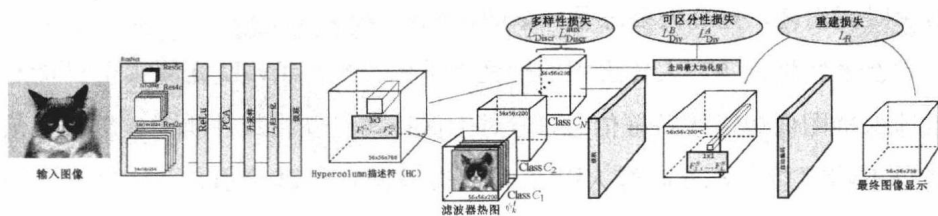


图 1.7 AnchorNet 架构

对语义匹配的研究不仅可以在像对上进行操作，还可以扩展到多幅图像中。多图像语义匹配可以找到多个图像间的一致对应关系，从而在应用中发挥更为重要的作用，如对象类模型重建^[26]和自动地标标注^[27]。Wang 等人^[28]将多图像间的语义匹配问题转化为特征选择与标注问题，即从每幅图像的初始候选集中选择一组稀疏特征点，通过分配标签建立它们在图像间的对应关系。该方法可以满足循环一致性和几何一致性的图像集合建立可靠的特征对应关系，其中循环一致性可以对图像集中的可重复特征进行选择和匹配。低秩约束用于确保特征对应的几何一致性，并可同时对循环一致性和几何一致性进行优化。该方法具有高度可扩展性，可以对数千幅图像进