

面向图像匹配的局部特征提取研究进展

刘向增,徐雪灵,刘如意,宋建锋,苗启广

(西安电子科技大学 计算机科学与技术学院,陕西 西安 710071)

摘要:图像匹配作为计算机视觉领域的重要研究方向,广泛应用于图像配准、图像融合、变化检测、视觉导航、3D重建、视觉同时定位与地图构建(SLAM)等领域,精确稳健的局部特征提取是实现其高效处理的前提与关键。以图像匹配研究为导向,从传统特征设计到现代特征学习对局部特征提取方法进行了分类总结,首先,为增强对现代局部特征提取方法的理解,重点介绍了基于传统特征设计的相关方法,接着回顾了基于经典机器学习的方法,搭建起传统方法到深度学习方法的桥梁,最后详细讨论了基于深度学习的现代特征提取方法。针对跨传感器、多视角、不同时段环境下的图像匹配需求,全面分析了各阶段主流方法的优缺点,提出了目前存在的问题与挑战,并给出了相应的研究建议,为相关研究人员全面深入理解图像局部特征提取方法并利用深度学习方法对其进行改进提供基础性参考。

关键词:图像匹配;局部特征;特征描述;不变特征;深度学习

中图分类号:TP311.5

文献标识码:A

文章编号:1673-629X(2022)02-0001-13

doi:10.3969/j.issn.1673-629X.2022.02.001

Research Progress of Local Feature Extraction for Image Matching

LIU Xiang-zeng, XU Xue-ling, LIU Ru-yi, SONG Jian-feng, MIAO Qi-guang

(School of Computer Science and Technology, Xidian University, Xi'an 710071, China)

Abstract: As an important research direction in the field of computer vision, image matching is widely used in image registration, image fusion, change detection, visual navigation, 3D reconstruction, simultaneous visual positioning and map construction (SLAM) and other fields. Among them, accurate and robust local feature extraction is the prerequisite and a key step to achieve efficient feature matching. Guided by image matching research, the local feature extraction methods are classified and summarized from traditional feature design to modern feature learning. Firstly, in order to enhance the understanding of modern local feature extraction methods, we focus on the related methods based on traditional feature design, then review the methods based on classic machine learning, build a bridge from traditional methods to deep learning methods, and finally discuss in detail based on modern feature extraction methods for deep learning. In view of the requirements of image matching under cross-sensor, multi-view, and different time periods, the advantages and disadvantages of mainstream methods at each stage are comprehensively analyzed, and the current problems and challenges of local feature extraction in image matching are put forward, and corresponding research recommendations are proposed. It provides a basic reference for related researchers to fully understand the topic of local feature extraction and improve the results with deep learning.

Key words: image matching; local feature; feature descriptor; invariant feature; deep learning

0 引言

随着智能系统与传感技术的发展,图像匹配技术已广泛应用于图像配准、多模式图像融合、变化监测、机器人视觉导航、目标跟踪、3D场景重建等领域。图像匹配精度易受数据获取环境(光照、天气)、平台姿态(视角、位置、运动状态)、传感器成像机理差异(红外、可见光、雷达)的影响(见图1),而实现图像精确、

稳健匹配的关键与前提是效局部特征的提取与描述。所谓局部特征是图像中不同于其邻域的一种模式,该模式与图像的某种特性变化相关联,如灰度、颜色、纹理、结构等^[1]。局部特征的表征形式可为显著点、直线、边缘,亦可为局部区域块、深层处理(梯度统计、滤波、卷积)的特征向量描述,因而,图像局部特征可分为语义性特征和定位性特征两类^[2];

收稿日期:2021-04-11

修回日期:2021-08-12

基金项目:国家重点研发计划(B019030051);国家自然科学基金(61771369);陕西省自然科学基金(2020JM-195,2020JQ-330);西安市科技计划项目(GXYD6.2);中央高校基本科研业务费(20101216855)

作者简介:刘向增(1981-),男,博士,副教授,硕导,CCF会员(D1746M),研究方向为计算机视觉与图像处理。

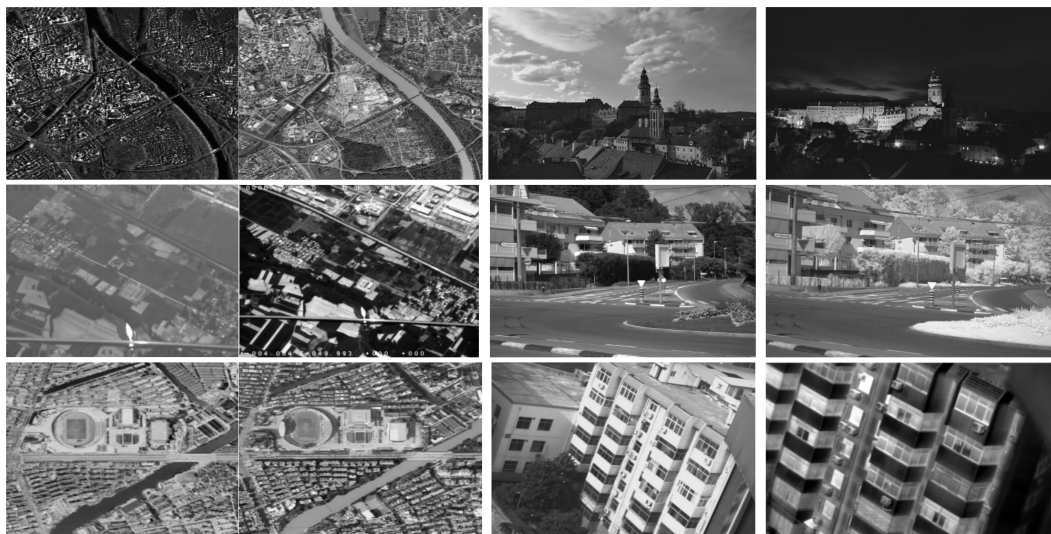


图 1 同一场景的多模态图像

(1) 语义性特征:具备某种目标属性的局部特征,如遥感图像中的直线特征时常对应道路,边缘对应图像纹理的分界线,图像的区域块特征对应某种特定纹理植被、水域或建筑。

(2) 定位性特征:不具备特定的目标属性含义,但其具备时空几何稳定性,常用于图像目标定位与特征追踪。

鉴于上述定义,语义性特征常用于目标检测与识别、图像检索、目标跟踪;定位性特征常用于图像匹配^[3]、配准与拼接^[4]、视觉导航^[5]、3D 重建^[6]、SLAM^[7]等领域。图像匹配领域中数据获取的环境(天气、时间、状态)不同及传感器差异使得图像间存在较大的辐射变化及几何结构形变,导致一致性局部特征的提取与描述变得异常困难。

图像匹配应用中高效的局部特征应具有以下性质:

(1) 可重复性:同一场景的两幅待匹配的图像中提取的局部特征应达到较高的重复率才能满足特征匹配的高精度对应需求。

(2) 独特性与显著性:两幅待匹配图像中提取的局部特征应反映出图像中显著性的结构特征,且其位置与几何属性具备独特性以减少单幅图像中的自相关性,为提升特征的单应性匹配精度提供保障。

(3) 几何与对比度不变性:局部特征应具备对比度及几何不变性,即两幅图像间发生辐射变化及获取视角变化时,不变性为局部特征的一致性提供保证。

(4) 高效性:局部特征提取在具备上述性质的同时,尽量具有低复杂度运算,以保证实时任务中图像匹配的高时效性。

上述性质中可重复性与不变性是衡量高效局部特征的最重要的两个特性,是实现对比度变化与几何差异条件下稳健与精确图像匹配的根本保障。图像局部

特征在匹配领域发挥了举足轻重的作用,原因在于它不但解决了前景与背景的辨识问题,而且可以对目标的任意部分进行定位与多尺度分析,同时在局部区域可将复杂的几何变化进行简化,实现快速有效的特征追踪。

局部特征的研究可以追溯至 1954 年 Attneave 提出的形状信息会集中至高曲率的控制点。该领域的研究已近 70 年,难以详细介绍所有的工作进展,文中主要针对常用的典型算法与模型进行归类与分析,主要从传统特征检测与描述设计、经典机器学习的特征定位与生成、现代深度局部特征学习等三个方面与读者进行分享,为相关研究人员提供比较全面的参考。

文中第一节主要介绍基于传统特征设计的局部特征提取方法与基本流程,为后续基于学习的模型设计提供基础;接着第二节展示了经典机器学习的特征定位与生成,为复杂环境下局部特征的自动检测与生成提供新思路;结合当前流行的深度学习,第三节对深度局部特征的学习模型进行的简介与分析;随后,第四节针对不同应用对当前流行的算法模型进行了概述;最后,第五节给出了结论。

1 传统局部特征

深度学习兴起之前,传统局部特征设计在图像匹配领域占据着至关重要的位置,甚至当今在某些特定环境下,传统局部特征的匹配效果依然优于深度学习方法。本节通过局部特征在图像匹配的应用,介绍传统局部特征的发展及典型成果,为理解后续基于学习方法的局部特征提取提供理论性基础。通常图像匹配的流程可分为特征检测、描述与匹配三部分,具体实现流程如图 2 所示,因此,本节分别从特征检测、特征描述两方面对传统局部特征设计进行回顾与总结,并对相关典型算法及其应用进行讨论。

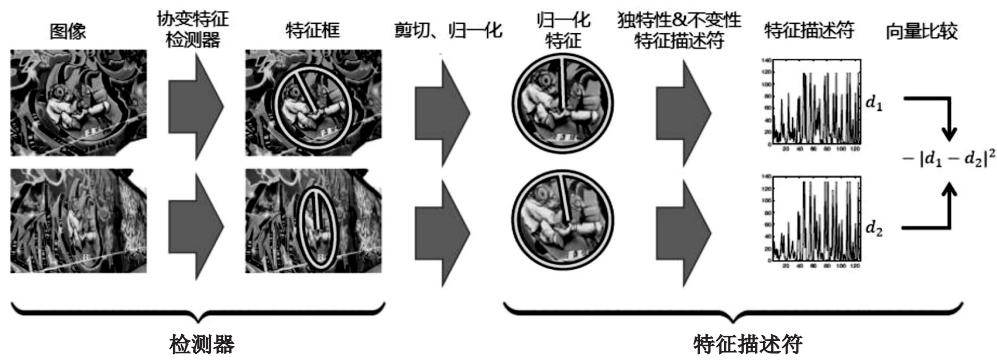


图 2 图像匹配中传统局部特征检测与描述

1.1 局部特征检测

传统局部特征检测是通过手工特征检测方法设计,利用特定的计算准则直接从图像中将特征进行定位与提取的过程。此类特征主要包括点特征、块特征、区域特征、线特征,典型的代表性算法如图 3 所示。

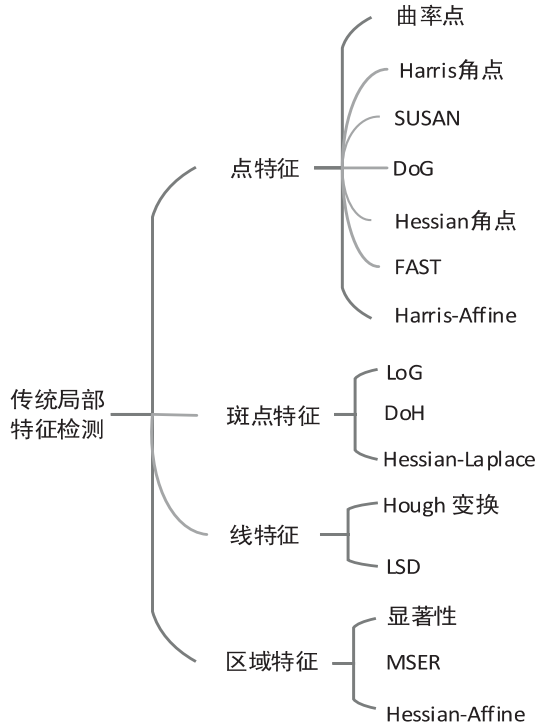


图 3 传统代表性局部特征检测方法

(1)作为最重要的局部特征,点特征是进行特征匹配的关键,是通过分析图像局部灰度的变化而提取的,可分为基于轮廓曲率、基于梯度变化、基于灰度分布三类。

轮廓曲率:目标轮廓的交叉点或结合点通常可反映信号的双向的变化,因此,利用轮廓曲率的极值可定位点特征^[8]。为精确定位曲率点的尺度信息,文献^[9]为应对尺度的变化提出了曲率图的尺度空间表示。Zhu 等^[10]提出了非线性点特征检测方法,建立了特征点检测的一系列标准,克服了曲率近似及高斯滤波的问题。

梯度变化:为改进 Moravec 提出的基于块自相关

的计算方法,Harris 等^[11]提出了通过分析图像的局部二阶矩阵(自相关矩阵),构建窗口平移信息量变化函数对角点特征精确检测的经典方法,具体函数如下:

$$M = \sum_w \begin{bmatrix} I_x^2(x,y) & I_x(x,y)I_y(x,y) \\ I_x(x,y)I_y(x,y) & I_y^2(x,y) \end{bmatrix} \quad (1)$$

其中, I_x, I_y 为 $I(x,y)$ 的偏导数。

$$F = \det(M) - \alpha \cdot \text{trace}(M) \quad (2)$$

其中, $\det(M)$ 与 $\text{trace}(M)$ 分别为矩阵 M 的行列式和迹, α 的取值范围为 0.04 ~ 0.06。为减少噪声的影响,通常提前对图像进行高斯滤波。

Harris 算子具有旋转与对比度不变性(见图 4),但缺乏尺度不变性。

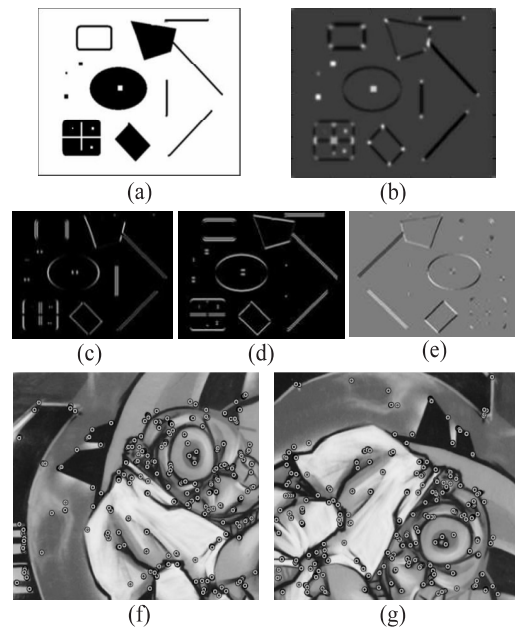


图 4 Harris 角点检测示例

注:(a)原图像;(b)角点检测结果;(c),(d),(e)分别为 $I_x^2, I_y^2, I_x \cdot I_y$;(f)和(g)为存在旋转的图像检测结果。

Mikolajczuk 与 Schmid^[12]将 Harris 算子融入高斯尺度空间实现特征检测的尺度不变性,二阶矩阵改为如下形式:

$$\mu = \sigma_D g(\sigma_I) \begin{bmatrix} L_x^2(x, \sigma_D) & L_x L_y(x, \sigma_D) \\ L_x L_y(x, \sigma_D) & L_y^2(x, \sigma_D) \end{bmatrix} \quad (3)$$

其中, L_x 为高斯平滑后微分图像, σ_D 为局部微分尺度, σ_I 为积分尺度。文章进一步考虑了各向异性的高斯仿射尺度空间, 实现了角点检测的仿射不变性。另外, Beaudet^[13] 利用 Hessian 矩阵(式 4) 行列式的最大值定位点特征。

$$H = \begin{bmatrix} I_{xx}(x, \sigma_D) & I_{xy}(x, \sigma_D) \\ I_{xy}(x, \sigma_D) & I_{yy}(x, \sigma_D) \end{bmatrix} \quad (4)$$

为实现点特征的多尺度检测并减轻计算复杂度, Lowe 提出的 SIFT 算法^[14] 中设计了 DoG (高斯差分) 算子(见图 5):

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (5)$$

$$L(x, y, \sigma) = G(x, y, k\sigma) * I(x, y) \quad (6)$$

通过搜寻 DoG 空间层与尺度层的极值点提取关键点, 然后利用边界效应去除和重定位, 实现点特征的亚像素定位。

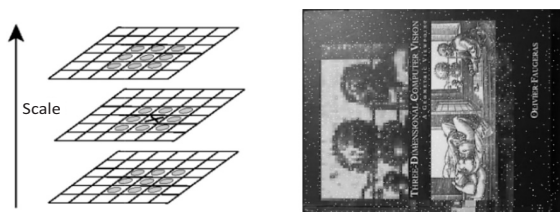


图 5 DoG 极值点检测

灰度值分布: 为满足点特征快速提取的要求, 文献^[15] 提出 SUSAN (最小核值相似区) 算子, 通过计算圆域内与中心点灰度值相近的数量, 若小于某一阈值, 则该中心点被检测为特征点。近来, 文献^[16] 提出了 FAST 算子, 通过统计固定半径圆域内灰度值大于或小于中心点的数量, 若大于某一阈值, 则将中心点作为特征点。

(2) 作为点特征的补充, 斑点特征可反映内容一致性局部区域, 通常为具有边界的近似圆斑。常见方法有高斯拉普拉斯尺度极值:

$$\Delta^2 = \frac{\partial^2 L}{\partial x^2} + \frac{\partial^2 L}{\partial y^2} \quad (7)$$

Hessian 矩阵(式 4) 行列式与迹同时极值法 (DoH), 为实现斑点特征的尺度不变性, 文献^[12] 提出了 Hessian-Laplace, 效果如图 6 所示。

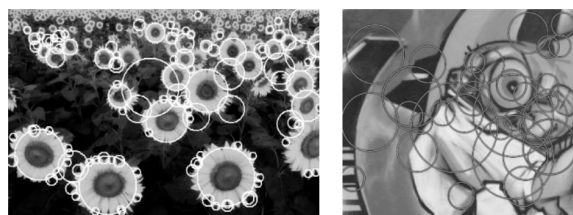


图 6 LoG 与 Hessian-Laplace 斑点检测

(3) 可反映图像结构性信息的另一类重要特征为线特征, 如 Hough 变换将图像中的直线检测问题转化为极坐标参数空间中通过某点最多正交曲线的问题^[17]。为提升直线检测的效率, Morel 等提出了 LSD^[18], 实现了图像中线特征的快速检测。

(4) 区域特征是反映局部纹理或内容一致性区域的特征, 较斑点特征的范围更广。常见的区域特征有显著性特征、最大稳定极值区域特征、仿射不变区域特征。Kadir 等将显著性特征定义为局部复杂性或不可预见性, 通过灰度概率分布信息熵的尺度空间极值确定特征的区域范围^[19]。作者进一步考虑尺度空间及椭圆区域形状参数实现了显著性区域的仿射不变性^[20]。Matas 等通过比较区域边界内外的灰度值实现最大稳定极值区域特征 (MSER) 检测^[21]。为实现仿射不变区域特征提取, 文献^[12] 提出了 Hessian-Affine 区域检测方法, 将 Hessian 矩阵的特征值作为椭圆区域的参数进行稳定性估计, 进而确定区域边界。

上述局部特征检测方法, 鉴于对比度及几何变化的影响, 点特征和区域特征应用比较广泛。为实现较好的匹配效果, 通常将多种特征进行组合检测。

上述局部特征检测方法, 鉴于对比度及几何变化的影响, 点特征和区域特征应用比较广泛。为实现较好的匹配效果, 通常将多种特征进行组合检测。

1.2 局部特征描述

局部特征提取后, 实现复杂环境下特征间稳健匹配的关键是如何对局部特征进行精确描述。特征描述子是对图像局部特征的定量化数学描述, 反映了其局部领域内的灰度、纹理或结构特性。理想的描述子应具备噪声、对比度、几何变化的不变性。特征描述方法主要分为梯度分布统计与局部二值化描述两类。

梯度分布统计描述: 该方法通过某种规则统计特征点邻域的梯度分布, 并生成描述向量。经典的方法如梯度方向直方图 (HoG)^[22], 对梯度的方向进行直方图统计, 如图 7 所示:

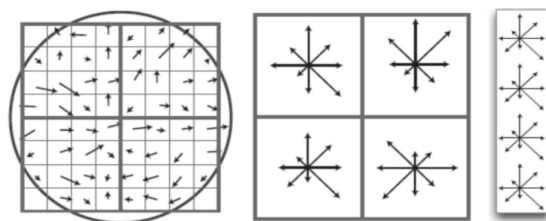


图 7 HoG: 梯度直方图统计描述

$$M(x, y) = \sqrt{L_x^2 - L_y^2} \quad (8)$$

$$\theta(x, y) = \arctan(L_y/L_x) \quad (9)$$

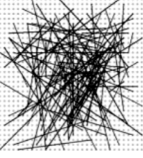
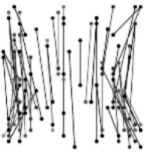
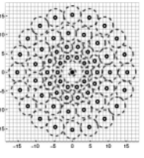
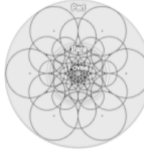
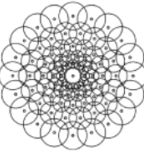
局部特征描述的里程碑工作^[23] 采用梯度直方图统计的思想对特征点进行 SIFT 向量描述, 实现同模态图像的稳健匹配。Bay 等提出了 SURF^[24], 利用 Haar 小波响应替代梯度并通过积分图像提升了描述子生成效率。Alcantarilla 将 SURF 算法扩展至非线性尺度空间, 提出 KAZE^[25] 进一步提升了描述子的匹配精度, 随后将快速显式扩散嵌入到金字塔框架, 实现了 KAZE 的加速计算^[26]。Abdel-Hakim 等将颜色信息融入描述子, 提出了 CSIFT^[27]。Morel 等通过生成不同

视角的图像并提取 ASIFT 向量,实现了全仿射不变性匹配同时增加了算法复杂度^[28]。Arandjelovi'c 等利用 Hellinger 核均方根替代欧氏距离,提出 RootSIFT 提升了特征匹配的效果^[29]。Dong 等提出了 DSP-SIFT^[30],通过特征块尺寸与梯度方向的池化,提升了描述子的匹配性能。Hassner 通过对特征点的多尺度

SIFT 子空间表示,提出 SLS^[31]实现了良好的匹配效果。

局部二值化描述:为满足移动与嵌入式设备的高效特征匹配及实时性任务(VSLAM)需求,研究人员提出了二值化特征描述(见表 1),利用 Hamming 距离对描述子进行快速匹配。

表 1 典型二值化描述子

	BRIFF	ORB	BRISK	FREAK	BIO
检测器		FAST-9, Harris 度量 尺度金字塔(1,5)	FAST-9, 尺度金字塔(4,3)		FAST-变形, DoH 度量, 尺度金字塔(1,8)
描述子					
方向		$M_{ij} = \sum_x \sum_y x^i y^j I(x,y)$ $c_x = \frac{M_{10}}{M_{00}}, c_y = \frac{M_{01}}{M_{00}}$ $C_{ori} = \tan^{-1}(c_y/c_x)$	$g(p_i, p_j) = \frac{p_j - p_i}{\ p_j - p_i\ } \cdot \frac{I(p_j, \sigma_j) - I(p_i, \sigma_i)}{\ p_j - p_i\ }$ $g = \begin{pmatrix} g_x \\ g_y \end{pmatrix} = \frac{1}{L} \cdot \sum_{(p_i, p_j) \in L} g(p_i, p_j)$	$\theta = \arctan \sum_{i=1}^N I_i - M \frac{dy_i}{dx_i}$ <p>s. t. $M = \text{median}(I_i)$ $i \in [1, \dots, N]$</p>	

轻量化与快速二进制检测器

拥有一个位编码框架(1:不同,0:相似)的二进制描述子

Ojala 等^[31]首次提出了局部二值模式(LBP),通过比较特征点与其邻域内像素的灰度值,若其值大则该位置为 1,反之为 0。为解决平坦区域描述子的不稳定问题,Heikkilä 等将 SIFT 中的梯度信息更替为 LBP,提出了 CS-LBP^[32]。随后,Chen 等通过改变编码的位数,提出了 RLBP^[33],并应用于纹理分类。Calonder 等于 2010 年利用特征点邻域的随机分布构建 BRIEF 描述子^[34],随后 Rublee 等通过 FAST 角点检测进行矩估计实现特征方向不变描述(ORB)^[35]。Leutenegger 等设计了围绕特征点的四个同心环采样模式,提出了 BRISK 二值特征描述方法^[36],通过尺度金字塔实现了尺度不变性。Alahi 等模仿人眼对中央凹区具备高分辨,对周边低分辨率的视觉感知的采样模式,提出了 FREAK 描述方法^[37]。受 LIOP^[38]对灰度信息进行排序的启发,Choi 等提出了灰度排序二值化特征描述(BIO)^[39],该方法有效缓解了灰度排序导致的误差。

传统局部特征描述主要对特征邻域内梯度或灰度分布进行简单统计,对同模式图像特征匹配效果良好,然而对于多模式图像及细节变化较大的图像依然效果欠佳甚至失效,因此,近来研究人员逐渐将学习的方法应用于局部特征的提取与描述。

2 经典学习型局部特征

基于经典学习方法的局部特征提取可归结为两

类:一是通过对传统局部特征描述的自适应学习,将特征投影至学习出的低维空间,实现局部特征的自动检测与描述;二是通过监督学习,依据特征的可匹配性与重复性构建特征分类器实现特征自适应提取。本节从学习型特征检测与描述两方面进行介绍。

2.1 学习型局部特征检测

最初,文献[25,45-46]利用经典学习方法进行加速传统局部特征检测。Hartmann 等^[40]训练随机森林分类器对 DoG 检测的特征进行可匹配性预测,进而提升 SFM 中匹配的效率。为抵抗天气、季节、时间引起的光照变化影响,Verdie 等提出了时不变学习检测器(TILDE)^[41],该方法将分段线性函数作为回归器获取的特征点在可靠性方面高于 SIFT 和 SURF,然而该方法不具备度不变性。

2.2 学习型局部特征描述

早期,Ke 等利用主成分分析(PCA)将梯度图像投影至新的空间构建 PCA-SIFT,实现了 SIFT 的降维处理,提升了匹配速度^[42]。随后, Lepetit 等利用随机树^[43]、Babenko 等^[44]通过 boosting 方法对匹配与非匹配特征进行表达学习。近来,Brown 等^[45]利用可操纵滤波或梯度方向映射对传统局部特征进行空间池化,并通过线性或非线性变换对特征进行降维(见图 8)。为应对由视角变化引起的仿射扭曲,Wang 等^[46]将 PCA 应用于输入图像块的仿射变换集合,通过线性空

间到点的映射获取特征的仿射子空间表达。对于较复杂的非刚体变换, Simo-Serra 等基于扩散几何^[47], 利用热扩散原理对图像块构建非刚体扭曲与光照不变描

述子 DaLI^[48], 利用 PCA 对特征降维保证描述子的紧凑性。

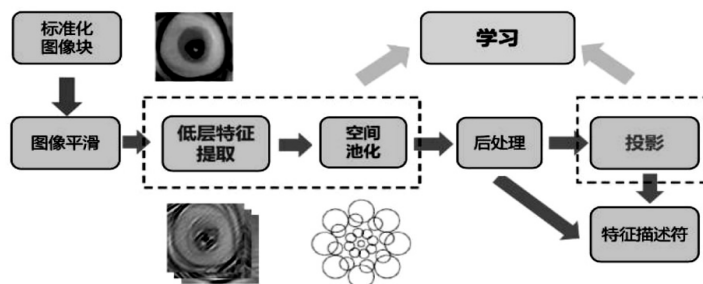


图 8 空间池化与特征嵌入优化的描述子学习

学习型二值化描述子的研究主要针对特定应用, 如人脸识别、运动恢复结构 (SFM)、3D 重建, 包括 LDAHash^[49], D-BRIEF^[50], RI-LBD^[51], BinBoost^[52], RFD^[53], BOLD^[54] 等算法。Strecha 等利用线性判别分析 (LDA) 最小化类内与类间距离比计算投影矩阵, 然后对阈值化投影生成二值化描述子 (LDAHash)。D-BRIEF 则通过学习图像块与判别子空间的线性投影映射对图像块进行二值化。

近来, Duan 等提出一种旋转不变二值化描述子的学习方法 (RI-LBD), 该方法先将图像块映射至旋转二值模式, 再学习块的方向与二值化投影矩阵。受 BRIEF 的启发, Trzcinski 等利用弱分类器对图像梯度进行池化, 通过提升二值化哈希函数进行二值化描述子生成。Fan 等提出了接受域描述子 (RFD), 依据区域的独特性或相关性进行的选取。Balntas 等通过对图像块的轻微扰动自适应调节度量, 进而生成二值化在线学习描述子 (BOLD), 其自适应性描述过程同样适用于其他二值化描述子。

基于经典学习的描述子利用学习的思想从特征匹配性、投影空间、几何扰动自适应等方面增加了传统局部特征提取与描述的灵活性, 在一定程度上提升了匹

配的效率。

3 深度学习局部特征

经典机器学习局部特征实现了传统局部特征的降维与自适应空间投影, 然而其本质并非自动学习的抽象特征。深度学习的兴起使得非线性自适应局部特征学习成为可能, 该方法利用多层级联非线性映射将输入图像直接进行特征提取, 不依赖传统特征设计。本节从深度学习特征检测、深度学习特征描述、一体化特征检测与描述三方面进行介绍。

3.1 深度学习局部特征检测

与传统局部特征提取方法类似, Lenc 等提出 CovDet^[55], 通过学习图像块与变换之间的映射将特征检测转化为回归问题。利用最小化损失函数使映射函数满足协变性约束:

$$\varphi(gx) = g\varphi(x), x \in X, g \in G \quad (10)$$

函数 φ 是由卷积神经网络 (CNN) 学习而得, 最终给出三种特征检测器, 前两种具备平移不变性, 第三种具备旋转不变性。作者将三种检测器应用于整幅图像将置信度超过某一阈值时作为特征点输出。作为 CovDet 的拓展, TcovDet^[56] (见图 9) 则将函数 φ 作为

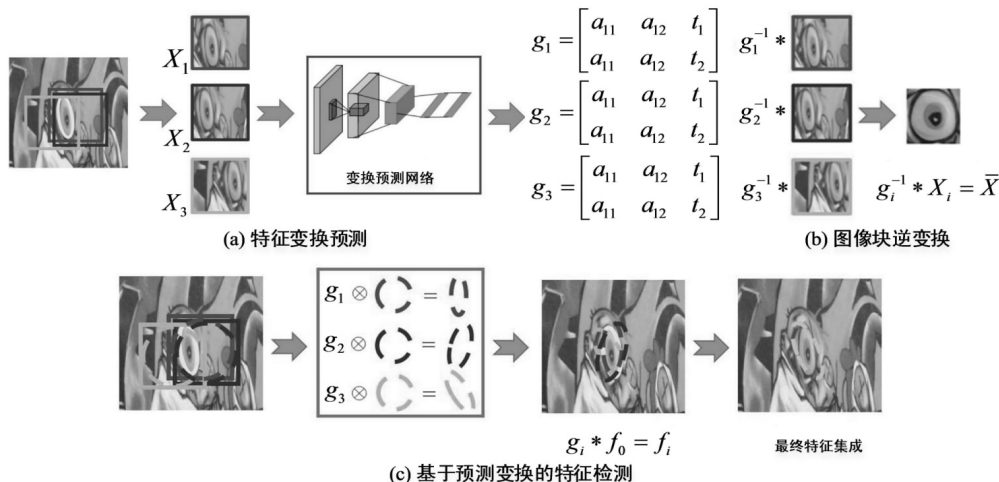


图 9 TcovDet 特征点检测流程

变换预测网络进行学习,将变换推广为仿射变换,实现了良好的特征检测。Savinov 等^[57]将特征检测转化为函数响应的学习问题,通过建立图像块与实数之间的映射并对其进行排序提取特征点。随后, Mishkin 等^[58]首次利用深度学习模型进行仿射协变区域检测,考虑可匹配性与几何精确性的平衡设计硬负常量损失函数:

$$L = \frac{1}{n} \sum_{i=1, n} \max(0, 1 + d(s_i, s_i) - d(s_i, N)) \quad (11)$$

其中, $d(s_i, s_i)$ 为描述子 s_i 与 s_i 之间的距离, $d(s_i, N)$ 为 s_i 与最难区分负样本之间的距离:

$$d(s_i, N) = \min_{i \neq j}(\min_{j \neq i}(d(s_i, s_j)), \min_{j \neq i}(d(s_j, s_i))) \quad (12)$$

该模型提取的仿射区域在图像检索方面获得了良好的匹配效果。Laguna 等^[59]提出浅层多尺度架构融合传统与学习的 CNN 滤波,通过传统对特征进行定位、排序,利用 CNN 滤波实现多尺度特征检测。

3.2 深度学习局部特征描述

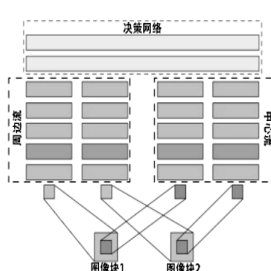
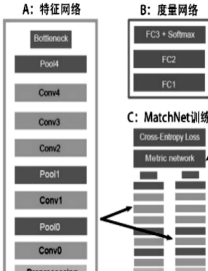
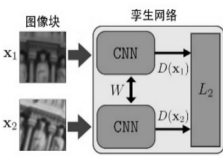
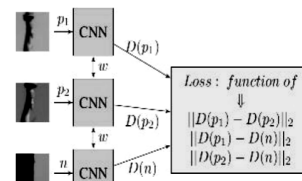
特征检测之后如何对其进行稳健的描述是后续特征匹配的关键。本节主要讨论基于深度学习模型的特征描述,从深度特征描述、度量学习、损失函数设计三

个方面进行介绍。

深度特征描述: AlexNet 在 ImageNet 的性能提升掀起了深度学习在视觉邻域研究的热潮,因此,早期研究人员直接将 AlexNet 提取的特征图作为图像描述子应用于图像匹配。Fischer 等最早将 AlexNet 的中间响应层作为特征,以 MSER 区域为描述范围构建描述子学习模型 PhilippNet 并应用于 3D 特征点之间的匹配。随后,Paulin 直接以 AlexNet 为深度特征提取模型,构建块卷积核网络(patch-CNN)^[60]进行快速的块特征描述。上述文献均以欧氏距离度量深度特征描述子的相似性,为应对图像间的复杂变化,研究人员对相似性度量进行了深入研究。

度量学习: 2015 年 Han 等^[61]将描述子与度量进行联合学习提出了 MatchNet 模型(见表 2),该模型通过多层 CNN 与空间池化进行描述子计算,利用三层全卷积层进行度量学习,结合交叉熵损失将匹配问题转化为分类问题。Zagoruyko 等^[62]提出了 Deepcompare,设计了孪生、伪孪生、双通道、空间池化孪生、双流多分辨等多种网络,并对其进行了比较,最终给出双通道网络性能最优。

表 2 基于孪生神经网络局部特征学习模型

	DeepCompare	MatchNet	DeepDesc	PN-Net
架构	描述子+匹配器 	描述子+匹配器 	描述子 	描述子 
方法	孪生网络+度量学习 伪孪生网络、孪生网络	孪生网络+度量学习 孪生网络	孪生网络 孪生网络	三元网络 三元网络
函数	合页损失	交叉熵损失	合页损失	SoftPN 损失
网络	$\min_w \frac{\lambda}{2} w_2 + \sum_{i=1}^N \max(0, 1 - y_i o_i^{\text{net}})$	$E = -\frac{1}{n} \sum_{i=1}^n [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)]$	$l(x_1, x_2) = \begin{cases} \ D(x_1) - D(x_2)\ _2, & p_1 = p_2 \\ \max(0, C - \ D(x_1) - D(x_2)\ _2), & p_1 \neq p_2 \end{cases}$	$l(T) = \left[\left(\frac{e^{\Delta(p_1, p_2)}}{e^{\min(\Delta(p_1, n), \Delta(p_2, n))} + e^{\Delta(p_1, p_2)}} \right)^2 + \left(\frac{e^{\min(\Delta(p_1, n), \Delta(p_2, n))}}{e^{\min(\Delta(p_1, n), \Delta(p_2, n))} + e^{\Delta(p_1, p_2)}} - 1 \right)^2 \right]$
输出	相似度分数	二值化标签	描述子(128 维)	描述子(256 维)

损失函数设计:为比较图像块之间的相似性,大多文献通过计算 CNN 描述子之间的距离进行判断,然而 CNN 特征由于不同的损失函数定义存在差异。本节主要从逐段损失、三元组损失、全局损失等进行讨论。Jahner 等首次将逐段损失应用于描述子的学习, Simo-

Serra 等提出 DeepDesc^[63]挖掘最难的正负样本提升网络的性能。

Choy 等针对几何对应提出了通用对应网络(UCN)^[64],利用卷积空间变换实现精确的几何对应,模型中采用的一致性对比损失如下:

$$L(x, x') = \begin{cases} \|f - f'\|_2 & x = x' \\ \max\{0, m - \|f - f'\|_2\} & x \neq x' \end{cases} \quad (13)$$

其中, m 为超参数。

Balntas 等提出利用三元组损失训练描述子匹配模型 PN-Net^[65], 其中 P, N 分别表示正负样本。Mishchu 等提出 HardNet^[66], 该模型通过简单的三元组损失最大化正样本与负样本之间的距离:

$$L = \frac{1}{n} \sum_{i=1}^n \max(0, 1 + \|a_i - p_i\|_2 - d_h) \quad (14)$$

$$d_h = \min\{\min_{p \in P} \|a_i - p\|_2, \min_{a \in A_i} \|a - p_i\|_2\} \quad (15)$$

TGLoss 等^[67]将三元组损失替换为全局损失, 在保证匹配块距离与非匹配块距离分布方差之和较小的情况下最大化匹配块距离与非匹配块距离分布均值之和。L2-Net^[68]设计的全局损失考虑了三个方面: 最小化匹配描述子之间的距离、最小化描述子内部的相关性、最大化描述子中间特征图的相似性。

3.3 一体化特征检测与描述

近来, 端到端一体化特征检测与描述模型成为局部特征学习的研究热点。Yi 等首次提出了端到端特征检测与描述模型 LIFT^[69], 该模型由三部分组成: 检测器、方向估计器、描述子生成器, 并由两个空间变换相关联。第一个空间变换实现检测器提取图像块由大及小的剪切; 第二个空间变换根据方向估计器对剪切后的图像进行旋转, 其中检测器采用 TILDE, 描述子则采用 DeepDesc 进行计算。

深度局部特征 (DELFL)^[70]利用 ResNet50^[71]提取深度特征, 基于视觉注意进行特征点筛选, 最后通过 PCA 实现描述子的降维, 成功应用于大规模的图像检索。DeTone 针对多视角问题提出了特征点检测与描述自监督训练模型 SuperPoint^[72], 对于输入图像可由单通道同时实现特征点定位与描述 (见图 10)。首先利用合成数据训练特征检测器, 然后通过单应性自适应生成伪特征点, 进而训练 SuperPoint 网络, 其中特征点检测采用交叉熵损失, 特征点描述采用合页损失。

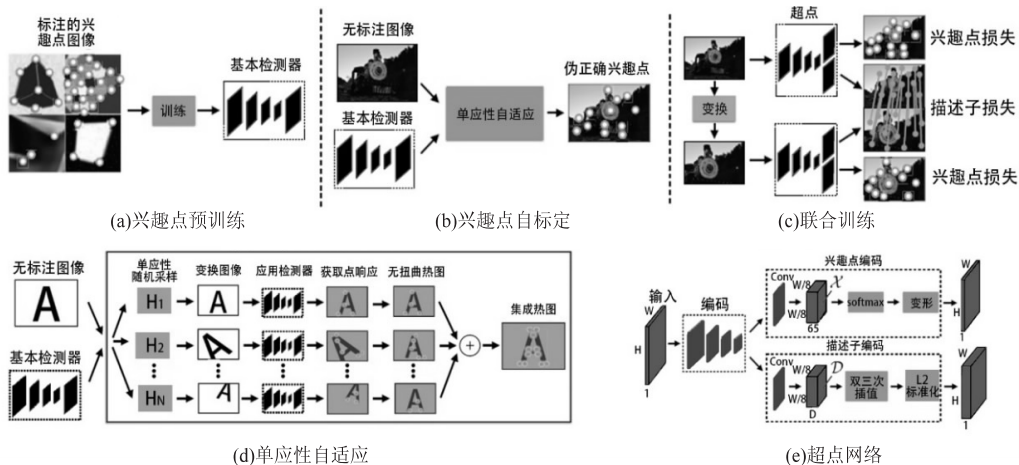


图 10 Superpoint 特征检测与描述自监督学习流程

Dusmanu 等用一个 CNN 实现特征检测与描述双重角色 (D2-Net)^[73], 其流程如图 11 所示, (i, j) 处的描述子由多层特征图组成, 特征点检测通过层内与层间软检测分值的非极大值抑制获得, 网络训练的损失函数采用特征检测与描述联合的三元组损失。Luo 等同时考虑视觉表征与特征之间的几何关系提出了 ContextDesc^[74], 该模型通过几何上下文编码器将特征点的几何关系进行编码, 利用视觉上下文编码器对局

部特征及区域特征进行编码, 最终构建出 128 维描述子。Sarlun 提出了基于注意力的上下文聚合机制 SuperGlue^[75] (见图 12), 首先通过关键点编码器将特征点位置及其描述子映射为同一个向量, 然后通过自注意和交叉注意层进一步转化为更为稳健的表征, 最后经过优化匹配给出分数矩阵并由 Sinkhorn 算法得到特征匹配关系。该模型利用 SuperPoint+SuperGlue 实现了室内与室外数据的良好匹配效果, 其性能超越了

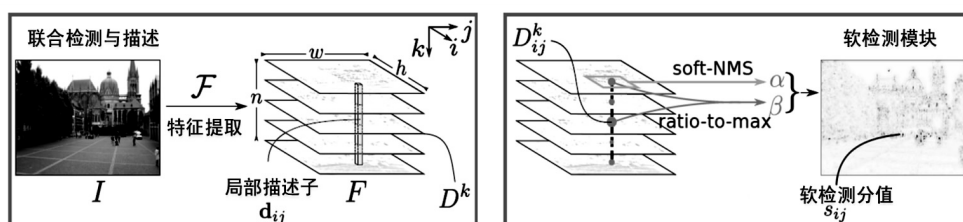


图 11 D2-Net 检测与描述流程

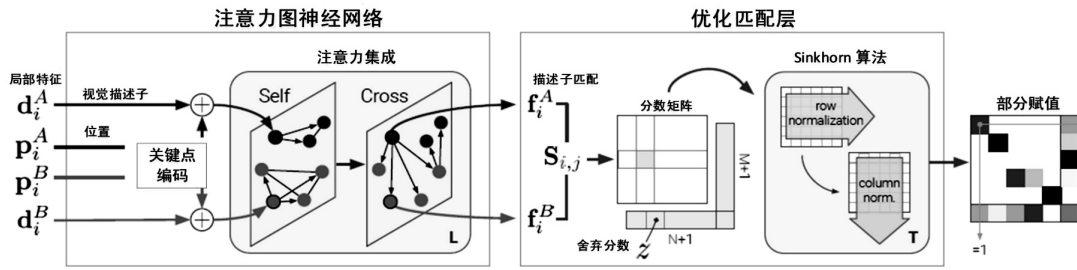


图 12 SuperGlue 实现架构图

D2-Net 与 ContextDesc。随后, Sun 等基于 SuperGlue 提出了由粗到细的匹配策略, 通过 Transformer 中的自注意与交叉注意层进行特征变换 (LoFTR)^[76], 再由置信矩阵得到初步对应关系; 细匹配阶段再利用 LoFTR 进行特征变换, 最后由相关系数矩阵给出匹配关系。

上述基于深度学习的局部特征提取与描述方法, 通过不同的模型构建、损失函数设计、匹配策略研究实现了较为灵活的特征提取与描述, 部分模型的性能已经超越传统局部特征, 为不同应用需求提供了诸多选择。

4 局部特征性能分析与比较

本节主要讨论传统局部特征、深度学习局部特征的性能比较, 为不同应用提供思路。

4.1 传统局部特征比较

早期, Heinly 等^[77]对多种传统特征检测与描述方法 (Harris, MSER, FAST, BRIEF, ORB, BRISK, SURF, SIFT) 在牛津数据集 (模糊、光照、视角等变化) 上从回忆率、重复率、正确匹配率等方面进行了分析与比较。当图像间不存在几何变化时, BRIEF 的性能优于 ORB, BRISK, SIFT; 当图像间存在旋转变换时, ORB 的性能优于 BRIEF 和 BRISK; 当图像间存在复杂的几何变化时, SIFT 的表现最优。

Mishkin 等^[78]提出了范围更广的宽基线立体匹配数据集, 包含几何、光照、传感器、外观、多模态等变化, 通过对多种检测器及描述子进行测试, 实验结果表明

Hessian-affine 具备最优的检测性能, SIFT 及其改进 DAISY 具有最佳的描述子表现。Sun 等^[79]提出了基于图像定位评价的购物中心场景数据集, 对 BRIEF, SURF, SIFT, COV^[80], RootSIFT 等局部特征提取方法进行了比较, 通过 COV 检测与 RootSIFT 描述获得了最佳性能。

4.2 深度学习局部特征比较

近来, 随着深度学习在局部特征提取的应用, 研究人员对学习方法与传统方法进行了应用性能的比较, 如文献[56, 81-82]。

Zhang 等^[56]将传统特征检测方法 with FAST、TILDE、CovDet、TCovDet 在多种数据集上进行了分析比较, 结果表明 TCovDet 在特征重复率方面优于 CovDet 和 TILDE, 这两种方法均优于 SIFT, SURF, MSER, Harris Laplace, Hessian affine, FAST 等检测器。在匹配性能方面, TCovDet 与 SIFT 分别在不同的数据集上获得了最佳表现。为解决在不同的数据集上对多种检测与描述方法的不一致问题, Balntas 等^[82-83]提出了新的基准数据集 HPatches, 该数据集包含大量适合描述子训练与测试的新数据, 并清晰定义了匹配、检索、分类等方面的评价协议。作者对 SIFT、RSIFT (RootSIFT)、BRIEF、ORB、BBoost、Deep-Compare (DC)、DeepDesc、TFeat^[83]、L2-Net、HardNet (HNet) 等进行了全面评价 (见图 13), 结果表明 L2-Net 与 HNet 在不同任务上获得了最优的性能表现, TFeat 在块证实与匹配方面优于传统局部特征检测与描述方法。此

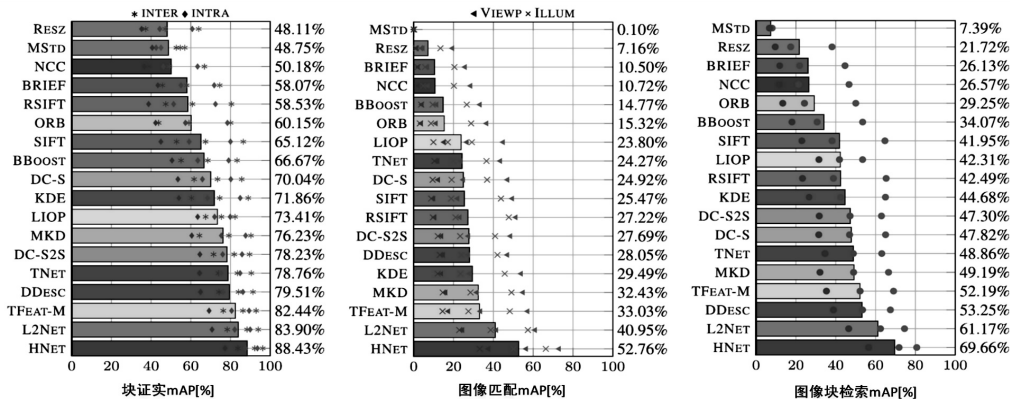


图 13 HPatches 数据集上不同方法在证实、匹配、检索等任务的精度比较

外, Lenc 等^[84] 基于 HPatches 数据集专门针对特征检测方法进行了评价, 提出了改进的特征检测重复率计算方法(见图 14), 在 5 类数据集上对 11 种检测方法进行测试, 统计结果显示 TILDE 有较高的特征检测重

复率, 然而在视角变化情况下 Hessian affine 的重复率达到最高, 某些情况下 TCovDet 的特征重复率依然最优。

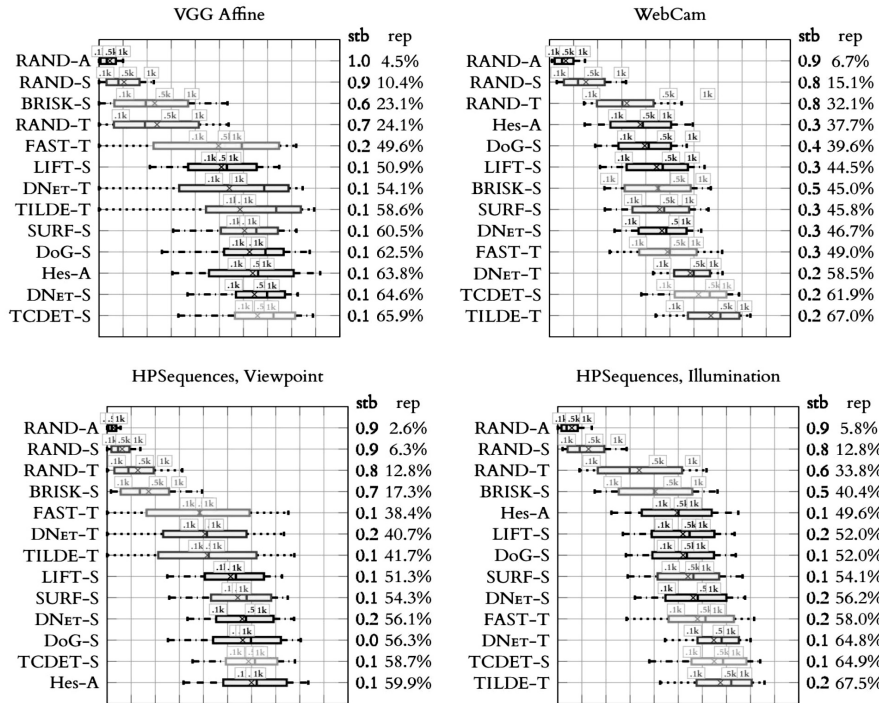


图 14 HPatches 数据集上不同特征检测方法的重复率比较

文献[88]对最新的基于深度学习的一体化特征检测与描述方法(D2Net, R2D2^[85], DISK^[86], Sparse-NCNet^[87], DRCNet^[88], LoFTR)进行了分析与比较, 结果表明 LoFTR 在单应矩阵估计及姿态估计的性能最优。

4.3 局部特征提取方法应用建议

通过前两节的比较与分析, 对比较具体的应用给出如下建议:

(1) 一般情况下传统局部特征提取方法较基于学习的方法计算复杂度低, 简单场景下特征检测可采用 Harris、FAST、DoG 等检测器, 特征描述可采用 SURF、BRIF、ORB 等描述子;

(2) 含有较大对比度、几何变换建议采用 SIFT、Hessian affine 等传统局部特征提取方法, 亦可采用 TILDE、TCovDet、SuperPoint 等特征检测方法;

(3) 复杂场景下的局部特征描述与匹配建议采用 RootSIFT、L2-Net、D2Net、ContextDesc、SuperGlue、LoFTR 等方法。

在较复杂的情况下(不同模态、不同时相、不同视角等), 也可将多种方法组合使用。

5 结束语

该文针对图像匹配的应用对局部特征的研究进展

进行了总结与分析, 主要从传统局部特征设计、经典机器学习局部特征提取模型构建、深度学习局部特征挖掘等三个方面对局部特征的检测与描述展开介绍与讨论, 分析了典型代表算法的基本原理及优势, 通过多篇文献的比较结果对具体的应用给出了相应的建议。目前, 局部特征提取仍存在诸多挑战, 如多模态、多时相导致的局部特征不一致; 多视角造成的几何变换过大, 局部特征不稳定。因此, 如何利用深度学习领域最新研究成果, 构建跨模态、多时相、宽动态条件下稳健的局部特征提取模型, 实现图像匹配、视觉导航、SLAM 等领域的更广范围的应用是今后研究的主要方向。

参考文献:

- [1] TUYTELAARS T, MIKOLAJCZYK K. Local invariant feature detectors; a survey[J]. Foundations & Trends in Computer Graphics & Vision, 2008, 3(3): 177-280.
- [2] CSURKA G, DANCE C R, HUMENBERGER M. From handcrafted to deep local features[J]. arXiv: 1807. 10254, 2018.
- [3] MA J, JIANG X, FAN A, et al. Image matching from handcrafted to deep features; a survey[J]. International Journal of Computer Vision, 2021, 129(1): 23-79.
- [4] PAUL S, PATI U C. Remote sensing optical image registration using modified uniform robust SIFT[J]. IEEE Geosci-

- ence and Remote Sensing Letters,2016,13(9):1300–1304.
- [5] 成坚炼,韩军,张国强.基于半直接法视觉里程计的单目视觉惯性系统[J].微电子学与计算机,2021,38(4):35–39.
- [6] 郑秋梅,温阳,王风华.基于注意力机制和可分离卷积的双目立体匹配算法[J].微电子学与计算机,2021,38(5):42–47.
- [7] 刘建明,何晴,陈辉.基于改进Census变换与颜色梯度融合的立体匹配[J].微电子学与计算机,2021,38(9):38–44.
- [8] MOKHTARIAN F, SUOMELA R. Robust image corner detection through curvature scale space[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1998,20(12):1376–1381.
- [9] LANGRIDGE D J. Curve encoding and the detection of discontinuities[J]. Computer Graphics and Image Processing,1982,20(1):58–71.
- [10] ZHU P, CHIRLIAN P M. On critical point detection of digital shapes[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,1995,17(8):737–748.
- [11] HARRIS C, STEPHENS M. A combined corner and edge detector[C]//Alvey vision conference. New York:[s. n.],1988:147–151.
- [12] MIKOLAJCZYK K, SCHMID C. Scale & affine invariant interest point detectors[J]. International Journal of Computer Vision,2004,60(1):63–86.
- [13] BEAUDET P. Rotational invariant image operators[C]//4th international conference on pattern recognition. Tokyo:[s. n.],1978:579–583.
- [14] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision,2004,60(2):91–110.
- [15] SMITH S M, BRADY J M. SUSAN—a new approach to low level image processing[J]. International Journal of Computer Vision,1997,23(1):45–78.
- [16] ROSTEN E, DRUMMOND T. Machine learning for high-speed corner detection[C]//European conference on computer vision. Berlin, Heidelberg:[s. n.],2006:430–443.
- [17] DUDA R O, HART P E. Use of the Hough transformation to detect lines and curves in pictures[J]. Communications of the ACM,1972,15(1):11–15.
- [18] VON GIOI R G, JAKUBOWICZ J, MOREL J M, et al. LSD:a fast line segment detector with a false detection control[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2010,32(4):722–732.
- [19] KADIR T, BRADY M. Saliency, scale and image description[J]. International Journal of Computer Vision,2001,45(2):83–105.
- [20] KADIR T, ZISSERMAN A, BRADY M. An affine invariant salient region detector[C]//European conference on computer vision. Prague, Czech Republic; Springer,2004:228–241.
- [21] MATAS J, CHUM O, URBAN M, et al. Robust wide-baseline stereo from maximally stable extremal regions[J]. Image and Vision Computing,2004,22(10):761–767.
- [22] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//IEEE computer society conference on computer vision and pattern recognition. San Diego; IEEE,2005:886–893.
- [23] BAY H, ESS A, TUYTELAARS T, et al. Speeded-up robust features (SURF)[J]. Computer Vision and Image Understanding,2008,110(3):346–359.
- [24] ALCANTARILLA P F, BARTOLI A, DAVISON A J. KAZE features[C]//European conference on computer vision. Berlin, Heidelberg:[s. n.],2012:214–227.
- [25] ALCANTARILLA P F, SOLUTIONS T. Fast explicit diffusion for accelerated features in nonlinear scale spaces[C]//British machine vision conference. Bristol:[s. n.],2013:1–11.
- [26] ABDEL-HAKIM A E, FARAG A A. CSIFT;a SIFT descriptor with color invariant characteristics[C]//IEEE computer society conference on computer vision and pattern recognition. New York; IEEE,2006:1978–1983.
- [27] MOREL J M, YU G. ASIFT;a new framework for fully affine invariant image comparison[J]. SIAM Journal on Imaging Sciences,2009,2(2):438–469.
- [28] ARANDJELOVIC R, ZISSERMAN A. Three things everyone should know to improve object retrieval[C]//IEEE conference on computer vision and pattern recognition. Providence; IEEE,2012:2911–2918.
- [29] DONG J, SOATTO S. Domain-size pooling in local descriptors;DSP-SIFT[C]//IEEE conference on computer vision and pattern recognition. Boston; IEEE,2015:5097–5106.
- [30] HASSNER T, FILOSOFF S, MAYZELS V, et al. SIFTing through scales[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2016,39(7):1431–1443.
- [31] OJALA T, PIETIKAINEN M, HARWOOD D. A comparative study of texture measures with classification based on featured distributions[J]. Pattern Recognition,1996,29(1):51–59.
- [32] HEIKKILA M, PIETIKAINEN M, SCHMID C. Description of interest regions with local binary patterns[J]. Pattern Recognition,2009,42(3):425–436.
- [33] CHEN J, KELLOKUMPU V, ZHAO G, et al. RLBP:robust local binary pattern[C]//British machine vision conference. Bristol:[s. n.],2013:1–10.
- [34] CALONDER M, LEPETIT V, STRECHA C, et al. Brief:binary robust independent elementary features[C]//European conference on computer vision. Berlin, Heidelberg:[s. n.],2010:778–792.
- [35] RUBLEE E, RABAUDE V, KONOLIGE K, et al. ORB:an efficient alternative to SIFT or SURF[C]//International conference on computer vision. Barcelona; IEEE,2011:2564–

- 2571.
- [36] LEUTENEGGER S, CHLI M, SIEGWART R Y. BRISK; binary robust invariant scalable keypoints [C]//International conference on computer vision. Barcelona: IEEE, 2011: 2548–2555.
- [37] ALAHI A, ORTIZ R, VANDERGHEYNST P. Freak; fast retina keypoint [C]//IEEE conference on computer vision and pattern recognition. Providence: IEEE, 2012: 510–517.
- [38] WANG Z, FAN B, WU F. Local intensity order pattern for feature description [C]//International conference on computer vision. Barcelona: IEEE, 2011: 603–610.
- [39] CHOI Y, PARK C, LEE J Y, et al. Robust binary feature using the intensity order [C]//Asian conference on computer vision. Singapore: [s. n.], 2014: 569–584.
- [40] HARTMANN W, HAVLENA M, Schindler K. Predicting matchability [C]//IEEE conference on computer vision and pattern recognition. Columbus: IEEE, 2014: 9–16.
- [41] VERDIE Y, YI K, FUA P, et al. Tilde; a temporally invariant learned detector [C]//IEEE conference on computer vision and pattern recognition. Boston: IEEE, 2015: 5279–5288.
- [42] KE Y, SUKTHANKAR R. PCA-SIFT; a more distinctive representation for local image descriptors [C]//IEEE computer society conference on computer vision and pattern recognition. Washington: IEEE, 2004: 506–513.
- [43] LEPETIT V, FUA P. Keypoint recognition using randomized trees [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(9): 1465–1479.
- [44] BABENKO B, DOLLAR P, BELONGIE S. Task specific local region matching [C]//IEEE 11th international conference on computer vision. Rio de Janeiro: IEEE, 2007: 1–8.
- [45] BROWN M, HUA G, WINDER S. Discriminative learning of local image descriptors [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 33(1): 43–57.
- [46] WANG Z, FAN B, WU F. Affine subspace representation for feature description [C]//European conference on computer vision. Zurich: [s. n.], 2014: 94–108.
- [47] GEBAL K, BARENTZEN J A, AANAS H, et al. Shape analysis using the auto diffusion function [J]. Computer Graphics Forum, 2009, 28(5): 1405–1413.
- [48] SIMO-SERRA E, TORRAS C, MORENO-NOGUER F. Dalli; deformation and light invariant descriptor [J]. International Journal of Computer Vision, 2015, 115(2): 136–154.
- [49] STRECHA C, BRONSTEIN A, BRONSTEIN M, et al. LDA-Hash; improved matching with smaller descriptors [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 34(1): 66–78.
- [50] TRZCINSKI T, LEPETIT V. Efficient discriminative projections for compact binary descriptors [C]//European conference on computer vision. Berlin, Heidelberg: [s. n.], 2012: 228–242.
- [51] DUAN Y, LU J, FENG J, et al. Learning rotation-invariant local binary descriptor [J]. IEEE Transactions on Image Processing, 2017, 26(8): 3636–3651.
- [52] TRZCINSKI T, CHRISTOUDIAS M, LEPETIT V. Learning image descriptors with boosting [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 37(3): 597–610.
- [53] FAN B, KONG Q, TRZCINSKI T, et al. Receptive fields selection for binary feature description [J]. IEEE Transactions on Image Processing, 2014, 23(6): 2583–2595.
- [54] BALNTAS V, TANG L, MIKOLAJCZYK K. Bold-binary online learned descriptor for efficient image matching [C]//IEEE conference on computer vision and pattern recognition. Boston: IEEE, 2015: 2367–2375.
- [55] LENC K, VEDALDI A. Learning covariant feature detectors [C]//European conference on computer vision. Amsterdam: [s. n.], 2016: 100–117.
- [56] ZHANG X, YU F X, KUMAR S, et al. Learning discriminative and transformation covariant local feature detectors [C]//IEEE international conference on computer vision. Venice: IEEE, 2017: 6818–6826.
- [57] SAVINOV N, SEKI A, LADICKY L, et al. Quad-networks; unsupervised learning to rank for interest point detection [C]//IEEE conference on computer vision and pattern recognition. Honolulu: IEEE, 2017: 1822–1830.
- [58] MISHKIN D, RADENOVIC F, MATAS J. Repeatability is not enough; learning affine regions via discriminability [C]//European conference on computer vision. Munich: [s. n.], 2018: 284–300.
- [59] BARROSO-LAGUNA A, RIBA E, PONSÁ D, et al. Key-net; keypoint detection by handcrafted and learned CNN filters [C]//IEEE/CVF international conference on computer vision. Seoul: IEEE, 2019: 5836–5844.
- [60] PAULIN M, DOUZE M, HARCHAOUÍ Z, et al. Local convolutional features with unsupervised training for image retrieval [C]//IEEE international conference on computer vision. Santiago: IEEE, 2015: 91–99.
- [61] HAN X, LEUNG T, JIA Y, et al. Matchnet; unifying feature and metric learning for patch-based matching [C]//IEEE conference on computer vision and pattern recognition. Boston: IEEE, 2015: 3279–3286.
- [62] ZAGORUYKO S, KOMODAKIS N. Learning to compare image patches via convolutional neural networks [C]//IEEE conference on computer vision and pattern recognition. Boston: IEEE, 2015: 4353–4361.
- [63] SIMO-SERRA E, TRULLS E, FERRAZ L, et al. Discriminative learning of deep convolutional feature point descriptors [C]//IEEE international conference on computer vision. Boston: IEEE, 2015: 118–126.
- [64] CHOY C B, GWAK J Y, SAVARESE S, et al. Universal correspondence network [C]//International conference on neural information processing systems. Barcelona: [s. n.],

- 2016;2414–2422.
- [65] BALNTAS V, JOHNS E, TANG L, et al. PN-Net: conjoined triple deep network for learning local image descriptors [J]. arXiv;1601.05030,2016.
- [66] MISHCHUK A, MISHKIN D, RADENOVIC F, et al. Working hard to know your neighbor’s margins: local descriptor learning loss [C]//International conference on neural information processing systems. Long Beach: [s. n.],2017;4827–4838.
- [67] KUMAR BG V, CARNEIRO G, REID I. Learning local image descriptors with deep siamese and triplet convolutional networks by minimising global loss functions [C]//IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE,2016;5385–5394.
- [68] TIAN Y, FAN B, WU F. L2-net: deep learning of discriminative patch descriptor in euclidean space [C]//IEEE conference on computer vision and pattern recognition. Honolulu: IEEE,2017;661–669.
- [69] YI K M, TRULLS E, LEPETIT V, et al. Lift: learned invariant feature transform [C]//European conference on computer vision. Amsterdam: [s. n.],2016;467–483.
- [70] NOH H, ARAUJO A, SIM J, et al. Large-scale image retrieval with attentive deep local features [C]//IEEE international conference on computer vision. Honolulu: IEEE,2017;3456–3465.
- [71] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE,2016;770–778.
- [72] DETONE D, MALISIEWICZ T, RABINOVICH A. Superpoint: self-supervised interest point detection and description [C]//IEEE conference on computer vision and pattern recognition workshops. Salt Lake City: IEEE,2018;224–236.
- [73] DUSMANU M, ROCCO I, PAJDLA T, et al. D2-net: a trainable CNN for joint description and detection of local features [C]//IEEE/CVF conference on computer vision and pattern recognition. Long Beach: IEEE,2019;8092–8101.
- [74] LUO Z, SHEN T, ZHOU L, et al. Contextdesc: local descriptor augmentation with cross-modality context [C]//IEEE/CVF conference on computer vision and pattern recognition. Long Beach: IEEE,2019;2527–2536.
- [75] SARLIN P E, DETONE D, MALISIEWICZ T, et al. Super-glue: learning feature matching with graph neural networks [C]//IEEE/CVF conference on computer vision and pattern recognition. Seattle: IEEE,2020;4938–4947.
- [76] SUN J, SHEN Z, WANG Y, et al. LoFTR: detector-free local feature matching with transformers [C]//IEEE/CVF conference on computer vision and pattern recognition. Nashville: IEEE,2021;8922–8931.
- [77] HEINLY J, DUNN E, FRAHM J M. Comparative evaluation of binary features [C]//European conference on computer vision. Berlin, Heidelberg: [s. n.],2012;759–773.
- [78] MISHKIN D, MATAS J, PERDOCH M, et al. Wxbs: wide baseline stereo generalizations [C]//British machine vision conference. Swansea: [s. n.],2015;1–11.
- [79] SUN X, XIE Y, LUO P, et al. A dataset for benchmarking image-based localization [C]//IEEE conference on computer vision and pattern recognition. Honolulu: IEEE,2017;7436–7444.
- [80] PERDOCH M, CHUM O, MATAS J. Efficient representation of local geometry for large scale object retrieval [C]//IEEE conference on computer vision and pattern recognition. Miami: IEEE,2009;9–16.
- [81] BALNTAS V, LENC K, VEDALDI A, et al. HPatches: a benchmark and evaluation of handcrafted and learned local descriptors [C]//IEEE conference on computer vision and pattern recognition. Honolulu: IEEE,2017;5173–5182.
- [82] BALNTAS V, LENC K, VEDALDI A, et al. H-Patches: a benchmark and evaluation of handcrafted and learned local descriptors [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2020,42(11);2825–2841.
- [83] BALNTAS V, RIBA E, PONSA D, et al. Learning local feature descriptors with triplets and shallow convolutional neural networks [C]//British machine vision conference. New York: [s. n.],2016;3–14.
- [84] LENC K, VEDALDI A. Large scale evaluation of local image feature detectors on homography datasets [J]. arXiv;1807.07939,2018.
- [85] REVAUD J, WEINZAEPFEL P, DE SOUZA C, et al. R2D2: repeatable and reliable detector and descriptor [C]//Conference on neural information processing systems. Vancouver: [s. n.],2019.
- [86] TYSZKIEWICZ M J, FUA P, TRULLS E. DISK: learning local features with policy gradient [C]//NIPSF/conference on neural information processing systems. [s. l.]; NIPSF,2020.
- [87] ROCCO I, ARANDJELOVIC R, SIVIC J. Efficient neighbourhood consensus networks via submanifold sparse convolutions [C]//European conference on computer vision. Glasgow: [s. n.],2020;605–621.
- [88] LI X, HAN K, LI S, et al. Dual-resolution correspondence networks [C]//NIPSF/conference on neural information processing systems. [s. l.]; NIPSF,2020.