DOI:10.13878/j.cnki.jnuist.2020.03.001

唐灿[1]　唐亮贵[1]　刘波[1]

图像特征检测与匹配方法研究综述

摘要

几十年来,图像特征检测与匹配一直是图像处理的最核心领域之一,是计算机视觉的基石.没有特征检测与匹配就没有 SLAM、Sfm、AR、通用图像检索、图像配准、全景图像等视觉任务.本文在回顾几十年来的经典检测算法的基础上,阐述了引用最新的以深度学习为首的机器学习算法后,在本领域取得的最新进展,包括特征点、局部特征子、全局特征子、匹配及优化、端到端框架等所有关键点,展示了算法各自的优缺点.总而言之,面对工业界的宽基线、实时、低算力检测的要求,图像特征检测和匹配仍然是一项未能完整攻克的任务,融合特征点、局部特征子、全局特征子、匹配及优化的多任务全局框架成为未来发展的趋势.

关键词

图像特征检测;描述子;匹配算法;深度学习

中图分类号 TP13

文献标志码 A

收稿日期 2020-01-16

资助项目 重庆工商大学开放课题(KFJJ2019067);重庆市教委课题(1792079)

作者简介

唐灿,男,副教授,主要研究方向为机器视觉.tangcan2003@126.com

1　重庆工商大学　计算机科学与信息工程学院,重庆,400067

0　引言

自 20 世纪 70 年代以来,图像特征检测与匹配一直是图像处理最核心领域之一,是计算机视觉的基石.人眼或相机接收到的是平面的二维图像,重建三维、理解世界、掌握世界一直是这个领域不懈的追求.从 20 世纪对纹理、颜色的理解到 21 世纪对线、点、面的特征提取,研究者们使用数学工具对这一过程进行了长达几十年的研究,取得长足的进展.最近 10 年,由于计算机算力、海量数据的快速增长,以深度学习为首的人工智能算法在计算机图像领域取得了丰硕成果,对图像特征检测与匹配领域产生了深刻的影响.新的研究表明:图像特征检测与匹配正在全面转向深度学习,从手工选择特征子转变为从数据中学习特征.但研究也同样表明:同时利用传统、可解释的检测匹配算法理论有助于更好地解决特征检测与匹配问题,有助于领域的进步和革新.

1　图像特征检测与匹配基本流程

一般而言,图像特征检测与匹配的核心流程如图 1 所示.

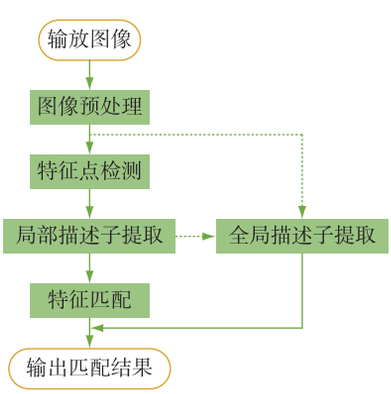

图 1　图像特征检测与匹配流程

Fig. 1　Flow chart for feature detection and matching

1)图像预处理:在进行特征检测之前,通常需要对图像进行预先的处理.包括灰度化、去噪声、生成图像金字塔等过程.不同的算法要求的预处理过程不一致.

2)特征点检测:提取图像中感兴趣的点,这些点被称为特征点,

这是图像稀疏化的一个典型过程.我们需要选择一些具有代表性的点来代理图像,检测出这些代表性的过程被称为特征点检测.三维重建依赖于这些点,所以特征点检测必不可少.

3)局部描述子(local descriptor)提取:通常,我们可以从特征点周围提取出一个小的几何区域(patch),并生成一个标识性的向量来代表这个区域的特征,这个特征向量被称为局部描述子或局部描述符.它将自己与其他区域区分开来,因而通常作为后续匹配过程的基础.

4)全局描述子(global descriptor)提取:用于描述整幅图像的全局特征向量.它代表着图像中的高层特征或语意,通常用于图像检索领域.全局描述子可以抽象自局部描述子,也可能是直接从图像中生成特征.

5)特征匹配:一旦有了局部描述子或全局描述子,就可以进行两个图像之间的匹配.找出图像间的匹配点,然后就可能利用 PnP(Perspective-n-Point)[1]、光束平差法(bundle adjustment)[2]进行三维重建等后续工作.

大多数算法可能只依赖于局部描述子或全局描述子的其中之一,因而无须同时生成两者.近来年,有少量算法同时生成两个描述子,并彼此依赖.

2 传统图像特征检测与匹配

传统的图像特征检测与匹配依赖精心挑选的手工检测算法,有着较为扎实的数学理论基础.

2.1 特征点检测

2.1.1 角点检测

角点检测是最早提出的特征点检测之一.角点没有严格的定义,但通常被视为两条边的交点,更狭义上讲,角点的局部邻域应该具有两个不同区域的不同方向的边界.在现实世界中,角点对应于物体的拐角,道路的十字路口、丁字路口等.从几何的角度上讲,角点通常表现为两个边缘的角上的点或邻域内具有两个主方向的特征点.角点是优秀的特征点,无论视角如何变换,这些点依然存在且稳定,并与邻域的点差别较大.但在实际应用中,大多数所谓的角点检测方法检测的是拥有特定特征的图像点,而不仅仅是"角点".这些特征点在图像中有具体的坐标,并具有某些数学特征,如局部最大或最小灰度、某些梯度特征等.

1)Moravec 角点检测算法[3]

Moravec 将"角"定义为自我相似程度低的点.因而,算法考虑以像素为中心点的一片范围,查看该范围与周围的相似度,如果相似度高,则不会被认为是"角",而在那些与附近像素的周围图像都很不相似的像素,才会被认为是"角".相似度通常是将两个范围的对应点计算误差的平方和,其值越小代表相似度越高.

假设现在是对一个二维灰阶图像 I 来做检测.考虑选取一个固定像素点 (x,y) 为中心点,其周围像素为区块(patch),其中某点的位移为 (u,v),因此中心点向量 (x,y) 与 patch 所有点的差的平方和记为 (u,v),而对于每个 (u,v) 做不同加权,就可以得到:

$$E(u,v)=\sum_{x,y} w(x,y)[I(x+u,y+v)-I(x,y)]^2, \tag{1}$$

其中 $w(x,y)$ 为权值,在标准的 Moravec 算法中为常数 1.

Moravec 角点检测算法可以找出整个图像的局部最大值(局部最不相似的点),这些局部最大值就很有可能是我们想要检测到的"角".从这个意义上讲,Moravec 算法不算是严格意义上的角点算法.

2)Harris 算法[4]

Harris 算法建立在 Moravec 算子的基础之上,它对 Moravec 进行了严格的数学建模和改进.主要体现为:克服 Moravec 只检测 45°倍角的缺点,使用泰勒展开,覆盖所有方向的检测.

对于式(1)中的平方项进行泰勒展开式,假设 I_x 和 I_y 是 I 的偏微分,可以得到:

$$E(u,v)\approx [u \quad v]\boldsymbol{M}(x,y)\begin{bmatrix}u\\v\end{bmatrix}. \tag{2}$$

E 值容易受到噪声的干扰,因为窗口是二值方形窗口.Harris 改用具有平滑效果的高斯圆形窗口进行处理,减少了噪声的影响.

对泰勒展开后的结果矩阵进一步优化,无需进行矩阵的特征分解,只需估计矩阵的行列式和迹,即可以判断角点.

Shi 和 Tomasi 进一步优化了此算法,提出了 Shi-Tomasi 角点检测算法[5],Harris 角点检测算法的稳定性和域值中的 k 值有关,而 k 是个经验值,不好设定最佳值.Shi-Tomasi 角点检测假设一般图像每个像素所给出的函数值通常是光滑且稳定的,角点的稳定性其实和矩阵 $\boldsymbol{M}$ 的较小特征值有关,于是直接用较小的那个特征值作为分数,这样就不需要调整 k 值了.

从本质上讲,Harris 算法、Shi-Tomasi 算法都是基于梯度的检测算法,基于梯度的检测方法计算复杂度高,其图像中的噪声可以阻碍梯度计算.

3) FAST 算法[6]

事实上,上述的角点检测算法都显得太过学术,在工程化过程中面临着计算量较大、速度较慢的严重问题.对此,Rosten 等[6]以更加简单的方式来定义角点,并提出了一个快速而简洁的检测算法(FAST).

FAST 角点定义为:若某像素点与其周围领域内足够多的像素点处于不同的区域,则该像素点可能为角点.具体算法步骤如下:

①在图片中选择一个像素点 P,并把它的亮度值设为 I_p;

②以该像素点为中心作一个半径等于 3 像素的离散化的 Bresenham 圆,这个圆的边界上有 16 个像素,如图 2 所示;

③设定一个合适的阈值 t,如果在这个大小为 16 个像素的圆上有 n 个连续的像素点,它们的像素值要么都比 $I_p + t$ 大,要么都比 $I_p - t$ 小,那么它就是一个角点.

事实上,FAST 算法中的 N 值很难直接给出,所以文献[6]从 ID3 算法中学习了合适的 N 值,并采用非极大值抑制的方法解决从邻近的位置选取了多个特征点的问题.从这个意义上讲,FAST 算法已非严格意义上的数学算法,而是现代意义上的数据算法.

FAST 检测算法计算速度快,可以应用于实时场景中.在 FAST 特征提出之后,实时计算机视觉应用中特征提取性能才有显著改善.目前,FAST 算法以其高计算效率、高可重复性成为计算机视觉领域最流行的角点检测方法.

2.1.2 斑点检测

区别于角点和边缘,斑点(blob)是更具普通意义的特征点.斑点主要描述的是一个区域.该区域相对其周围的像素在颜色或者灰度上有明显区别.例如:从远处看,一颗树是一个斑点,一块草地是一个斑点,一个人也可以是一个斑点.由于斑点代表的是一个区域,相比单纯的角点,它的稳定性要好,抗噪声能力要强[7].

要检测出这样的“点”的思路也很简单,最直接的就是基于求导的微分方法.我们可以使用一阶微分算子或二阶微分算子求出这样的“点”,一个常用的考虑是使用拉普拉斯算子.拉普拉斯算子是简单的各向同性微分算子,它具有旋转不变性,所以可以方便的用于变化检测.但拉普拉斯算子对噪声比较敏感.1980 年,Marr 和 Hildreth 提出将拉普拉斯算子与高斯低通滤波相结合,提出了 LoG(Laplace and Guassian,高斯拉普拉斯算子)算子,从而大大降低了对噪声的敏感度[8].LoG 算子的缺点在于计算量大、处理速度慢.

1) DoG(Difference of Gaussian,高斯差分算子)与 SIFT 算法[9]

Lowe 于 1999 年提出了 SIFT 算法,并于 2004 年整理发表[9].SIFT 算法的全称为:尺度不变特征变换(Scale-Invariant Feature Transform,SIFT),它是一个完整意义上的解决方案,很大程度上解决了目标的旋转、缩放、平移、图像仿射/投影变换、光照影响、杂乱场景、噪声等重大难题.

由于计算机无法预知图像中物体的尺度,因而需要同时考虑图像在多个尺度下的描述,从而获知感兴趣物体的最佳尺度.如果某些关键点在不同的尺度下都相同,那么在不同尺度的输入图像下就都可以检测这些关键点匹配,也就是尺度不变性[10].

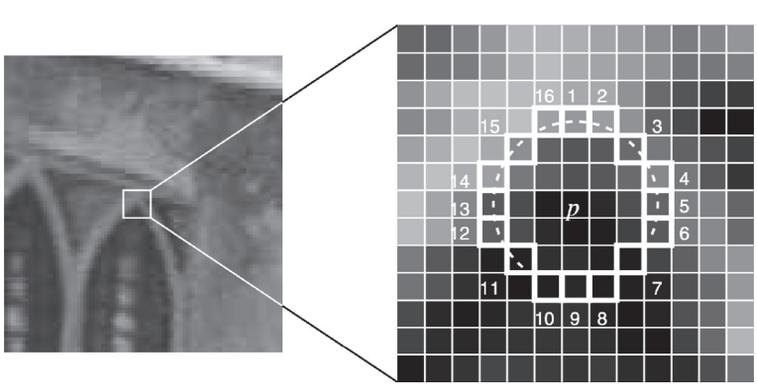

图 2 FAST 算法检测[6]

Fig. 2 Feature points detection for FAST[6]

SIFT 算法引入尺度空间理论,同时建议:在某一尺度上的特征检测可以通过对两个相邻高斯尺度空间的图像相减,得到 DoG 的响应值图像 $D(x,y,\sigma)$.然后仿照 LoG 方法,通过对响应值图像 $D(x,y,\sigma)$ 进行局部最大值搜索,在空间位置和尺度空间定位局部特征点将 LoG 算子简化为 DoG 算子.这样不仅可以得到更好的关键点,而且可以减少计算量.

SIFT 算法是近 20 年来传统图像特征检测算法中的标杆算法,具有里程碑意义,其谷歌学术的引用数高达 55 000 多次,通常用作特征检测算法的 Baseline 使用.与 SIFT 算法相比较事实上成为衡量一个算法优良程度的基准.由于其专利问题,所以在开源算法中使用不多.

2) SURF 算法[11]

SIFT 算法由于计算量巨大,不能用于实时系统中.2006 年,Bay 等改进了 SIFT 算法,提出了 SURF(Speeded-Up Robust Features,加速稳健特征)快速算法[11],在保持 SIFT 算法优良性能特点的基础上,解决了 SIFT 计算复杂度高、耗时长的缺点,提升了算法的执行效率.为了实现尺度不变性的特征点检测与匹配,SURF 算法先利用 Hessian 矩阵确定候选点,然后再进行非极大抑制.同时,为提高算法运行速度,在精度影响很小的情况下,用近似的盒状滤波器代替高斯核,并引用查表积分图,从而实现比标准 SIFT 算法快 3 倍的运行速度.

3) KAZE(风)算法[12]

传统的 SIFT、SURF 等特征检测算法都是基于线性的高斯金字塔进行多尺度分解来消除噪声和提取显著特征点的.但高斯分解是牺牲了局部精度为代价的,容易造成边界模糊和细节丢失.非线性的尺度分解希望解决这种问题,由此,KAZE 算法的作者 Alcantarilla 等[12]提出使用非线性扩散滤波法,将图像亮度(L)在不同尺度上的变化视为某种形式的流函数(flow function)的散度(divergence).由于非线性微分方程没有解析解,一般通过数值分析的方法进行迭代求解.传统上采用显式差分格式的求解方法只能采用小步长,收敛缓慢.KAZE 中采用 AOS(Additive Operator Splitting)算法对结果进行收敛.

在 KAZE 算法的基础上,Alcantarilla 等在 2013 年进行改进,提出了 AKAZE 算法[13].AKAZE 是加速版 KAZE 特征,即 Accelerated KAZE Features.作者引入快速显示扩散数学框架 FED 来快速求解偏微分方程,FED 的引入让它比之前的 AOS 更快更准确.

2.2 局部描述子

正如人眼做图像匹配一样,事实上,我们不能将图像的点与点匹配起来,只能将图像中的一块与另外图像中的一块匹配起来.换言之,我们匹配的其实是图像的局部特征,因而,在特征点周围选择一块区域,用一些特征向量对其进行描述就变得理所当然,这就是局部描述子.

局部描述子的核心问题是不变性(鲁棒性)和可区分性.由于使用局部图像特征描述子的时候,通常是为了鲁棒地处理各种图像变换的情况.因此,在构建/设计特征描述子的时候,不变性问题就是首先需要考虑的问题.在宽基线匹配中,需要考虑特征描述子针对视角变化的不变性、对尺度变化的不变性、对旋转变化的不变性等特性;在形状识别和物体检索中,需要考虑特征描述子对形状的不变性.

传统的描述子都是基于数学的方法精心挑选得出.SIFT 描述子是其中的佼佼者.首先它利用关键点邻域像素的梯度方向的分布特性,为每个关键点指定方向参数,从而保证了特征点的旋转不变性以及尺度不变性.然后再统计以特征点为中心的局部区域梯度,生成 128 维梯度特征向量,再归一化特征向量,去除其光照的影响.通过以上步骤产生的特征点具有旋转不变、尺度不变以及光照不变等性能,如图 3 所示.

SIFT 描述子最大的问题在于计算量大、效率不高,不利于后面的特征点匹配.事实上,并不是所有维都在匹配中有着实质性的作用.因而可以用 PCA、LDA 等特征降维的方法来压缩特征描述子的维度.在此基础上,发展出一大批的改进算法,例如 SURF 算法[11]、PCA-SIFT 算法[14]、SSIFT 算法[15]等.

BRIEF(Binary Robust Independent Elementary Features,独立、可靠的二进制基础特征)算法把局部描述子的简化做到了极致[16].它无需计算类似于 SIFT 的复杂特征描述子,只生成一个二值串即可.首先,在特征点周围选择一个块,在块内通过一种特定的方法来采样,挑选出 n 个点对.然后对于每一个点对(p,q),比较这两个点的亮度值,如果 $I(p) > I(q)$ 则这个点对生成了二值串中一个的值为 1,否则为 0.所有 n 个点对,都进行比较之后,就得到了一个 n 位长的二进制串,通常 n 可以设置为 128、256 或 512.对于一个 $S \times S$ 的块,标准 BRIEF 算法的(p,q) 采样方式为:p 和 q 都符合$(0,S^2/25)$ 的高斯分布.BRIEF 算法简单、实时性较好,但无法支持大角度的旋转,因

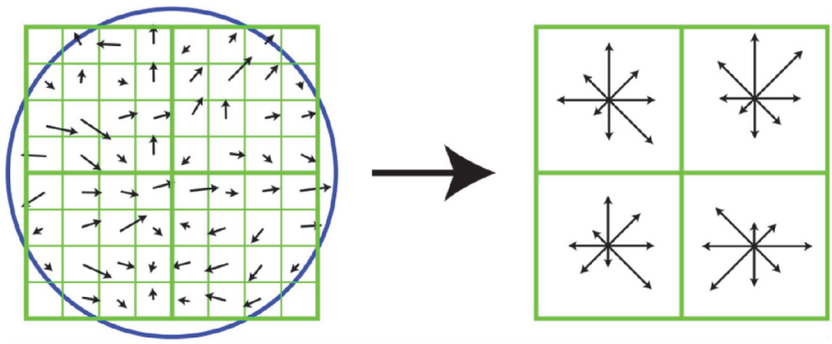

图 3 SIFT 局部描述子

Fig. 3 Local descriptor for SIFT algorithm

而需要增加其描述子的旋转不变性.

ORB(Oriented FAST and Rotated BRIEF,支持 FAST 方向和 BRIEF 旋转不变性)算法[17]来自于 Rublee 等的论文 *ORB: an efficient alternative to SIFT or SURF*,它是现今实时 SLAM 系统中应用最广泛的算法之一.其特征提取由 FAST 算法改进,利用图像金字塔为其增加了尺度不变性;特征点描述是根据 BRIEF 特征描述算法改进的,它利用灰度质心法计算方法来解决以及旋转不变性,并放弃手工选择的 n 对点,使用数据学习的方法来学习到如何选择 256 对点.事实上,传统手工算法从 2010 年后与学术算法之间的界限变得模糊,混合使用机器学习和手工特征子成为趋势.

2.3 特征点匹配

一旦有了特征描述子,我们就可以将图像的特征点两两对应起来,这个过程称为特征点匹配.特征点匹配最基本的方法是使用暴力匹配(Brute-force matcher),它将待匹配图片的特征描述子中每一行都与待匹配图片的描述子中每一行进行距离计算,从而得到最佳匹配.这个距离根据不同的描述子可能有不同的选择,比如 ORB 算法中使用汉明距离.暴力匹配最大的问题在于计算的时间复杂度和空间复杂度都比较高,因而引入 FLANN 匹配(Flann-based matcher),它使用快速近似最近邻搜索算法寻找,这是一种近似匹配,不一定能找到最佳匹配,但速度得到大大加快.优化的方法通常是使用索引,一般有线性索引、kd 树索引、k 均值索引、组合索引等.

事实上,直接使用描述子匹配总会遇到错误的匹配,这其中又通常分为两种:

1)假阳性匹配(False-positive matches):将非对应的特征点检测为匹配

2)假阴性匹配(False-negative matches):未将匹配的特征点检测出来

其中,假阳性匹配可以通用优化算法将其剔出.RANSAC(RANdom SAmple Consensus,随机抽样一致)算法[14]是使用最广泛的一致性优化算法.其核心思想就是随机性和假设性,它可以从一组包含“局外点”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,有一定的概率得出一个合理的结果,而为了提高概率必须提高迭代次数.随机性用于减少计算,循环次数是利用正确数据出现的概率.事实上,RANSAC 算法广泛用于各种一致性优化问题,未能考虑到图像优化匹配自身特点,如仿射变换、透视变换等.为此,相当多的算法对此进行了改进,引入仿射不变性的 ASIFT (Affine Scale Invariant Feature Transform) 算法[18]、引入透视不变性的 PSIFT[19] (Perspective Scale Invariant Feature Transform) 算法、变换一致性的 CODE[20](Coherence Based Decision Boundaries for Feature Correspondence)算法,这些算法都对匹配问题进行了各个方向的优化.

2017 年,Bian 等在 CODE 算法的基础上提出了一种简单快速的 GMS(Grid-based Motion Statistics,基于网格的运动统计)优化算法[21],它是一种基于网格的运动统计特性的方法,将运动平滑限制转成去除错误匹配的数据测量,使用一种有效的基于网格的分数估计方法,使得特征匹配算法能达到实时性.该方法可以迅速剔除错误的匹配,从而提高匹配的稳定性.图 4 展示了 SIFT 算法与 GMS 算法的对比效果.

3 基于学习的图像特征检测与匹配

手工算子总是基于这样或那样的前提假设,是对现实世界的简化和抽象.因而,在鲁棒性上和泛化

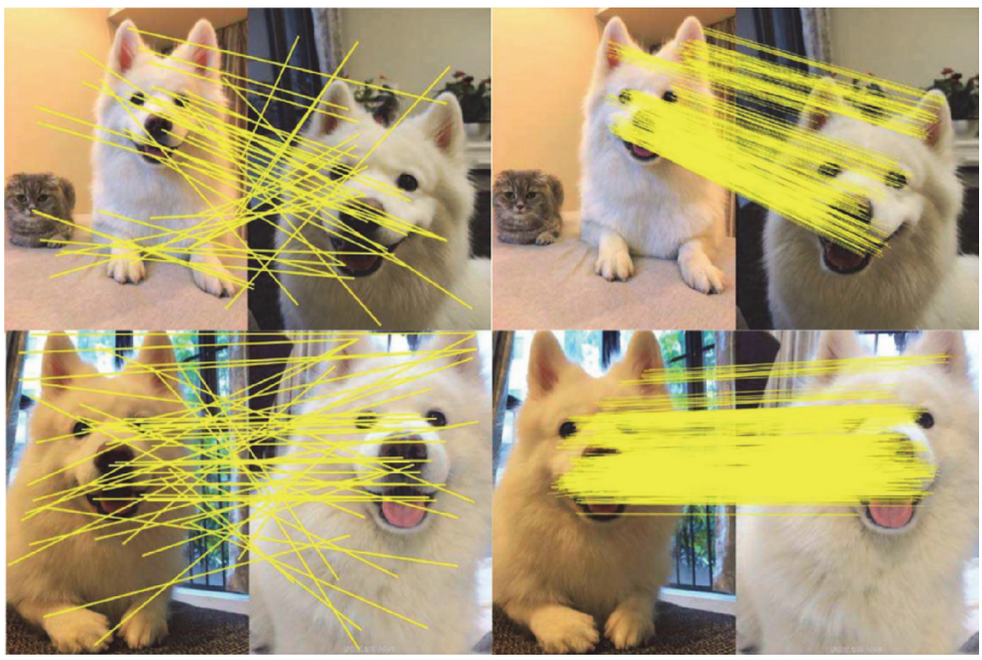

图 4　SIFT 算法匹配(左)和 GMS 匹配(右)[21]

Fig. 4　Matching result for SIFT (left),and GMS (right)[21]

能力方面有着天然的不足.近 10 年来,随着计算性能的不断攀升和大规模数据标注数据集的普及,以深度学习为首的机器学习算法成为研究和应用的基础,传统的手工标注的描述子逐渐向以数据驱动的学习算法转变.

3.1　基于学习的关键点检测

单独研究关键点检测算法并不多见,其原因在于:我们尝试对一张图像生成稀疏的兴趣点,但我们很难说明哪些是兴趣点,RGB 的兴趣点与深度图像的兴趣点是否一致.因而,一段时间以来,研究者很难提出取代传统算法的关键点检测算法.

Verdie 等提出了一个时间不变的学习探测器[22](Temporally Invariant Learned Detector,TILDE),用于解决在天气和光照条件急剧变化的情况下,检测可重复的关键点.它使用一组不同的季节、不同的时间,从相同的角度捕捉的相同场景下的训练图像,通过 DoG 来生成训练数据集,使用自定义的分段线性回归函数进行训练,并使用 PCA (Principal Component Analysis,主成分分析)进行优化,从而实现了在光照变化情况下,比 SIFT 更好的可重复特征点检测的性能.

学习算法依赖数据.如果使用关键点检测的学习,面临的另一个大的问题是:如何标注数据集? 如前所述,我们很难说明哪些点才是关键点,因而似乎无法进行人工标注,生成训练集.Quad-Networks 算法[23]使用无监督数据表达方式,训练神经网络以不变变换的方式对点进行排名,将学习兴趣点检测器的问题转化为学习排名点的问题.该算法认为兴趣点来自某些响应函数的顶部/底部分位数,因而从该排名的顶部/底部中提取兴趣点.

也有研究从已生成的众多手工特征点中进一步学习,提取出更加稳定的特征点.Key.Net[24]就是采用的这种方式,它从多个尺度上采用手工特征点,通过 CNN (Convolutional Neural Network,卷积神经网络)网络进行进一步的过滤,再复合到原尺寸的图像上,其网络结构如图 5 所示.

3.2　基于学习的局部描述子检测

局部描述子的作用在于提取局部图像的特征,通常我们可以把一整幅图像分成均等的块,每个小块被称为一个 patch.对于深度学习的研究者而言,所有的手工特征选择算子都远逊于类如 CNN 这样的特征提取网络,因而,使用 CNN 来取代传统的手工特征描述子成了自然而然的事.

文献[25]在这个方面进行了尝试.它把图像的 patch 成对地输入到 CNN 网络中,再加入决策网络用于判断其相似性.论文对网络结构进行了基本的尝试,从中选择孪生网络,实现了区分相似度的目的.文献[26]的思路与之类似,通过孪生 CNN 网络

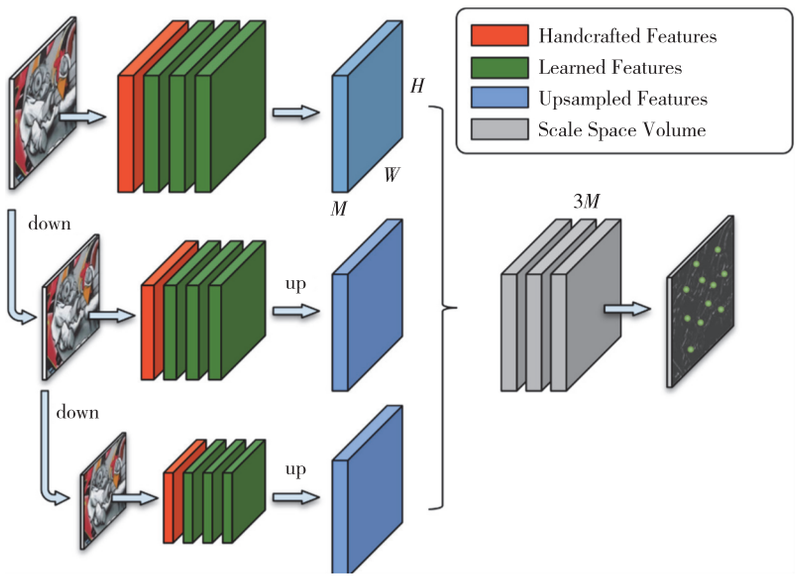

图 5　Key.Net 网络结构[24]

Fig. 5　Network structure diagram for Key.Net[24]

学习 patch 表示,在训练和测试期间,它使用 L2 距离,提出了一种 128-D 描述子,其欧几里德距离反映了 patch 相似性,用以替代 SIFT 的局部描述子.文献[27]则更进一步,但将网络分为特征提取网络、度量网络和度量训练网络三个部分,分别采用类 AlexNet 网络取特征、度量网络进行距离度量和孪生网络进行相似度判断.它没有采用传统的欧式距离,而是学习了一个三层全连接的度量网络.

如果说 2015 年时的深度网络还停留在相似匹配,2016 年后的网络则明显更进一步,学习到的特征使得让相同的更靠近,不同的更分离.文献[28]与文献[29]都不约而同地使用三元组损失函数进行训练,并且开始考虑算法的实用性,使用浅层神经网络进行特征提取.

L2-NET[30]提出了一个结构简单、特征提取效果较好的 CNN 网络,它提出递进的采样策略,可以保证网络在很少的 epochs 就可以访问到数十亿的训练样本,同时,重点关注 patch 特征之间的相对距离,也就是匹配上的 patch 对距离比未匹配上的 patch 对距离更近,从而取消了距离阈值的设置.此外,L2-NET 网络也包含了相对复杂的一个级联网络来处理中心块的信息,并在多个点设定了网络的多个损失函数,这也使得训练、收敛都相对困难.L2-NET 的网络结构如图 6 所示.

HardNet[31]在 L2-NET 的基础上进行了进一步的改进,它受到 Lowe 的 SIFT 的匹配标准启发,从困难样本入手,无需使用两个辅助损失项,只需要使用一个损失函数,简单却有效,学习到了更有力描述子,在图像匹配、检索、宽基线等方面都做了大量详细实验,在真实任务中取得了最先进的结果.作者在 Github 上提供了完整的代码,并使用多个数据集,不断提高其泛化能力.至此,图像特征提取有了一个可以真正进入实用阶段的算法.

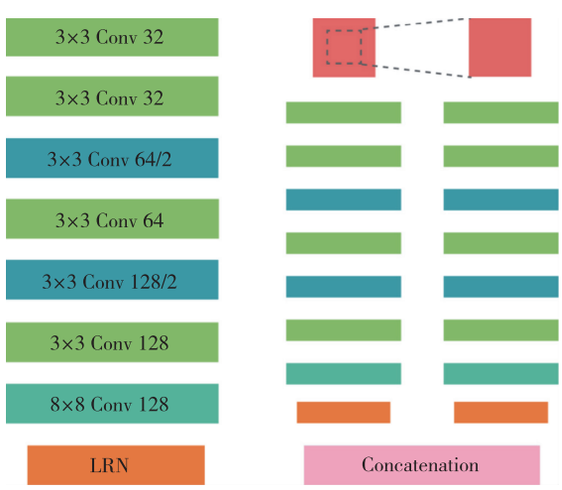

图 6　L2-Net 网络结构[30]

Fig. 6　CNN layers for L2-Net[30]

基于 CNN 的局部描述子学习,尽管在基于像素块的数据集上获得了很好的效果,却在 SFM 数据集上未能表现出良好的泛化性能.GeoDesc 算法[32]采用和 L2-Net 相似的网络结构,提出了一种融合了几

何约束的局部描述子学习方法,采用传统 SFM 方法,得到三维点及其对应的一系列像素块的对应关系.选用的像素块为 SFM 中使用的特征点,这样能够提高样本的准确性.算法整合了多视图重构中的几何约束关系,因而在数据生成、数据采样和损失函数的计算等方面促进了局部描述子的学习过程,生成的描述子被称为 GeoDesc 描述子.SOSNet[33]将二阶相似性(SOS)用于学习局部描述子,并提出了一个新的正则化项,称为二阶相似正则化(SOSR),通过将 SOSR 结合到训练中,在多个数据集上表现优秀.

事实上,无论是使用哪个数据集来训练网络,patch 块的选择都存在较大的问题,不同的描述子选择各不相同,通常必须由关键点检测器事先适当地估计其大小、形状、方向.如果两个补丁不对应,则它们的描述子将不匹配.为此,AffNet[34]探索影响学习和配准的因素:损失函数、描述子类型、几何参数化以及可匹配性和几何精度之间的权衡,并提出了一种新的硬负常数损失函数(Hard Negative-Constant Loss Function)用于仿射区域的学习.文献[35]则建议使用对数极坐标(log-polar)采样方案直接提取“支持区域”.通过同时对点的近邻进行过采样和对远离点的区域进行欠采样,可以提供更好的表示.它也证明了这种表示特别适合于使用深度网络学习描述子.此模型可以在比以前更大的比例尺范围内匹配描述子,还可以利用更大的支持区域而不会遭受遮挡.

3.3 基于学习的全局描述子检测

由于局部描述子专注细节、偏重纹理,因而,对于通用图像检索这类更加抽象的任务而言,使用局部描述子很难得到正确的检索结果,所以,需要一个更加高层次抽象的特征检测,即全局描述子检测.

从某种意义上讲,图像分类与目标检测也算是全局意义上的描述子,但它们受限于类别标签,无法提供通用检索.因而,对于通用图像检索,尤其是大规模图像检索而言,提取基于图像的全局描述子成为一个非常重要的选择.

BoW(Bag-of-Words)算法是简单直观的全局描述算法.BoW 算法源自文本分类领域的词袋模型.假定对于一个文档,忽略它的单词顺序和语法、句法等要素,将其仅仅看作是若干个词汇的集合,引入到图像领域后成为 BoVW(Bag of Visual Words)算法[36].它利用 SIFT、SURF 算法生成的局部描述子进行聚类,把最具代表的“单词”选择出来,构造成一个字典(codebook),用它来代表图像本身.聚类算法同样可以采用 K-means 聚类,或者随机森林(Random Forest)[37].FV(Fisher Vectors)编码算法[38]也被经常使用,它采用混合高斯模型(GMM)构建字典.不过,FV 不只是存储视觉词典在一幅图像中出现的频率,还统计了视觉词典与局部特征(如 SIFT)的差异.

VLAD(Vector of Local Aggregated Descriptors)系列算法是颇受关注的全局描述子算法.VLAD 算法[39]首先针对一张图像,提取了 N 个 D 维特征,再对全部的 $N\times D$ 特征图进行 K-means 聚类,获得 K 个聚类中心,接着获取并累加了每个聚类的所有特征残差,最终得到了 K 个全局特征.这 K 个全局特征表达了聚类范围内局部特征的某种分布,抹去了图像本身的特征分布差异,只保留了局部特征与聚类中心的分布差异,从而生成了特定大小的全局描述子,这样生成的编码也被称为 VLAD 编码.NetVLAD[40]在 VLAD 的基础上,使用 CNN 来进行全局描述子提取.将 VLAD 公式中的二值函数平滑化,转化为可微的函数算法.除此之外,它使用监督学习获得聚类中心,从而向真正把同一物体的类别聚在一起的目标跨进了一步.NetVLAD 最大的问题在于:输出特征的维度太大,使得无论是处理还是拟合都变得困难.NeXtVLAD[41]则在 NetVLAD 的基础上更进一步,它吸收了 ResNeXt 对 ResNet 网络进行改造的思想,在应用 NetVLAD 聚合之前,将其中的 FC 网络一分为三,将高维特征分解为一组相对低维的向量,从而达到了更强拟合并降维的目标.

除此之外,直接利用 CNN 从图像提取全局特征子的想法则更加普遍.Neural Code 描述子[42]首开先河,提出在大型分类数据集(如 Image-Net)上进行训练的分类卷积神经网络,靠近顶端的全连接层输出值可以直接用作图像视觉内容、语义级别的高层次描述子(descriptor).文献[43]引入 Sum-Pooled Convolutional features (SPoC) 取代 Max-Pooled Conv-features 作为图像全局描述子.文献[44]通过使用孪生网络和排序损失函数,改进 RMAC 描述子,将其投射为完全可区分的网络,从而产生了一个较好的图像描述子.文献[45]提出了一种 REMAP(Multi-Layer Entropy-Guided Pooling,多层熵引导池化)全局描述子.该描述子从多个 CNN 层中学习并聚集了深层特征的层次结构,并以三重态损失端对端地进行了训练.REMAP 明确学习了在视觉抽象的各种语义级别上相互支持和互补的判别特征,从而具有更好的代

表性.

3.4 基于学习的匹配算法

RANSAC 算法通常用于匹配后的优化算法,但只有一致性是远远不够,包括 GMS 算法在内的匹配算法总是提出类如分块平滑这样的简化假设,这些假设并不总是成立.使用学习算法引入真实数据,进行更加真实的匹配成为自然的选择.

文献[46]认为,通常的立体匹配图像之间存在类如本质矩阵(essential matrix)的约束,它是远比 RANSAC 算法更强的约束,因而应当利用这样的约束.文献[46]将生成的特征点视为点云,受 PointNET 的启发,通过深度网络来学习这种映射方式,拟合出点集的坐标对应关系到点的对应正确性的映射关系,从而实现更好的匹配.SuperGlue[47]是 Magic Leap 公司于 2019 年的最新成果,它使用图神经网络来匹配网络,通过共同查找对应关系并拒绝不可匹配的点来匹配两组局部特征.通过解决可微分的最优变换问题来估算分配,引入图神经网络预测其代价.SuperGlue 在上下文聚合机制加入注意力,使其能够学习 3D 世界的几何变换和规律性的先验知识,共同推理基本的 3D 场景和要素分配.其代码可以在现代 GPU 上实时执行匹配,并且很容易集成到现代 SFM 或 SLAM 系统中.

3.5 端到端(end-to-end)检测

深度学习优于传统算法的一个重要特点在于:可以设计一个网络,使得输入图像直接输出特征点、局部描述子.即:一个端到端的检测网络.

LIFT 算法[48]成为一个成功的起点.它设计了一个完整的深度网络体系结构,该体系结构实现了完整的特征点处理管道,即检测、方向估计和特征描述.在保持端到端的差异性的同时学习如何以统一的方式完成特征点检测、方向判断和局部描述子生成这三个问题,并证明:对这些单独的步骤进行优化并相互配合良好地运行有助于提升整体的检测性能.图 7 展示了 LIFT 图像的检测管道.

LF-Net[49]的思路与文献[48]类似,它使用一个检测网络生成一个尺度空间分数图和密集的方向估计,用于选择关键点位置、尺度和方向.用可微采样器(STN)对所选关键点周围的图像块进行裁剪,并将其反馈给描述子网络,每一个 patch 产生一个描述子.为训练网络,它设计了一个左右两分支的孪生结构,该结构以同一场景的两幅图像为输入.其深度图和摄像机的内、外特性都可以通过传统的 SFM 方法得到.再对右边的图像进行变形,以确定图像之间的 ground-truth 对应关系.最新的 RF-NET[50]是在 LF-NET 基础上改进的,它构造了更大的感受野特征图,从而导致更有效的关键点检测,同时引入一个广义的损失函数项——邻居掩码,以便于训练样本的选择,改进了稳定性.

SuperPoint[51]是 Magic Leap 公司于 2018 年发表的一篇文章,它基于自监督训练的特征点检测和描述子提取方法,是一个深度学习用于特征点提取和匹配的方法.它提出了一种可以自我学习的方法,通过构建 pseudo-ground truth 的特征点位置,并通过这些点本身来训练特征点检测器,从而避免了大量人力的手工标注.它首先手工生成一个具有简单几何形状特征的合成数据集,它们有着明确的特征点,再使用这些点来训练一个名为 MagicPoint 的检测网络,接着,引入 MS-COCO 数据集里未标注的图像,使用 MagicPoint 检测器来进行特征点检测并混合多个对应变换后的的图像特征点,这个过程相当于对图像进行标注.最后,结合特征点和描述子来训练基于 MagicPoint 网络的全卷积神经网络,于是就得到最终的检测器——SuperPoint.与 2016 年的 LIFT 相比,SuperPoint 优势明显,可用于 SLAM 中,对季节和环境光照具有更强的鲁棒性.

为了学到更加重复、稳定、可靠的特征,R2D2 算

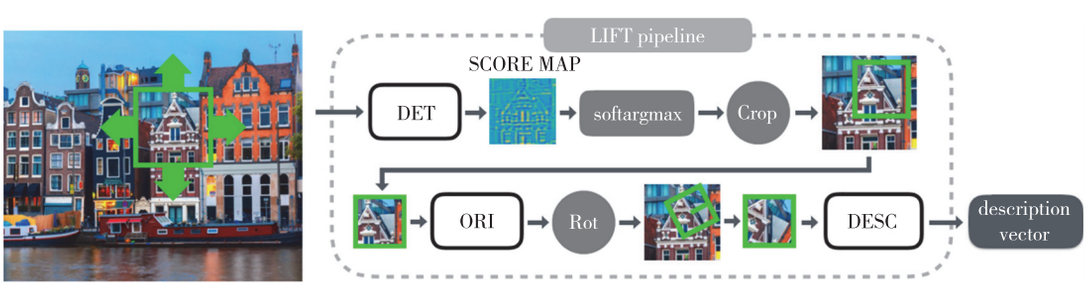

图 7 LIFT 图像检测管道[48]

Fig. 7 Pipeline for LIFT[48]

法[52]主张仅在可以高置信度执行匹配的区域中学习描述子.它同时学习关键点检测和描述子以及判断局部描述子的预测因子,从而避免出现歧义区域,生成可靠的关键点检测和描述子.

与传统的先生成特征点,再提取描述子不同,D2-Net[53]使用一个 CNN 网络,输入 $h\times w$ 的原始图片 I,输出 $f(I)$ 为一个 $h\times w\times c$ 的 3D 张量的特征图(Feature Map),可以将该特征图看作类似于 SIFT 等传统检测器中的 DoG 和特征子的混合,再从特征图里同时提取关键点和特征描述子(detect-and-describe).从某种意义上讲,特征描述子也就是关键点,关键点和描述子之间变成特征图的一体两面,思路颇有特色.

3.6 多任务融合匹配

事实上,现实中的任何一个图像匹配都不能仅仅使用局部描述子,而应当在更高的上下文和全局描述子的基础上进行匹配.如果再进一步,我们可以把图像特征点、局部描述子、全局描述子、匹配及优化多种任务合为一体,以框架的形式完整实现,才能真实地完成现实中的匹配.近年来的研究正在呈现这样的特点.

DELF(DEep Local Feature)[54]的架构是 Google 提出的一个以图搜图模式的图像检索架构.严格意义上讲它更加专注于生成全局描述子,然后进行高层匹配.它放弃了传统从局部描述子生成全局描述子的过程,而是直接使用图像级的类别进行弱监督学习得到的.为了识别图像检索中具有语义信息的局部特征,它还提出了一个关键点选择的机制,这个机理会共享更多网络层的描述子信息,对语义特征引入了注意力机制,因而表现力较强.生成描述子后,再使用 KD-tree 和 PQ 进行最近邻搜索,从而实现了快速查找的目标.2019 年,DELF 得以升级,使用 R-ASMK 算法[55]和最新的加入框的数据集,大大提高在地标方面进行数据检索的精度.但作为专一的架构,它只关心了图像检索任务,在 SFM 和 SLAM 上不具有通用性.

ContextDesc[56]通过引入上下文感知来扩展现有的局部特征描述子,从而超越了局部细节表示.它提出了一个统一的学习框架,该框架利用和聚合了跨模态上下文信息,包括来自高层图像表示的视觉上下文,和来自二维关键点分布的几何上下文,它融合所有的特征信息,加入一种预测“匹配能力”的度量,通过学习框架,实现更好的匹配的目标.

HF-Net[57]能够使用一个网络完成三项任务:生成全局图像描述子、检测特征点、局部特征点描述子.首先通过图像检索的方式(图像级的描述子检索)获取候选匹配“地点”,而后通过局部特征匹配实现精确的六自由度(6-DoF)位姿估计.由于图像全局描述子估计和局部特征的检测是两个分开的任务,如果采用两个网络,将会需要大量的计算量.为此,HF-Net 采用多任务学习的方法,通过两个任务共用部分网络,达到了通过一个网络同时估计全局描述子和提取局部特征的目的.该网络由一个共同的编码网络和三个“头部网络”组成.三个“头部网络”分别能输出全局的图像描述子、特征点检测响应图和特征点描述子.其中编码网络由一个 MobileNet 搭建而成,全局图像描述子由 NetVLAD 层输出,采用 SuperPoint 解码器实现特征点的提取和描述子的计算.为解决数据集难以满足的困难,网络还采用知识蒸馏的方法进行网络的训练,通过利用“教师网络”,减小了对数据集的要求.HF-Net 在网络效果上和实时性上都表现得较为突出.

4 总结

现实世界是丰富、复杂而多样化的,手工设计算法只是对现实世界的简化和抽象,因而很难适应宽基线下图像的可重复性、可区分性、准确性和高效性的要求,传统的手工设计的图像检测与匹配算法已近瓶颈.近十年来,以深度学习为首的图像检测与匹配正逐步走向主流.它们在所有的技术指标上都取得了或多或少的进步,更加难能可贵的在于:这一切都可能通过一个完整的 end-to-end(端到端)网络来加以实现.

但是,我们也要看到,这些算法仍然存在一些明显的缺陷.深度学习算法的问题之一在于数据集泛化问题,我们无法取得面对现有世界都具有普适性的数据集.因而,权重数据总会在这样或那样的场景下变得不那么可靠,这也是宽基线图像处理的最大问题之一.同时,深度学习的加入加重了计算的要求,一些算法需要大量的 CPU 甚至 GPU 的加入,在低功耗、实时要求较为苛刻的领域表现明显.

另一方面,随着计算能力的不断增强,嵌入式的前端变得越来越智能化.一些不具备简单视觉识别能力的设备将会更多地使用到传统的基于数学建模的检测算法,而一些有条件提供较强算力的设备则会充分使用深度学习算法和机器学习算法.我们正

处在一个由表形到表意变革的初始阶段,图像特征检测与匹配将在这里发挥出基础而关键的作用.

近年来,随着深度学习、图神经网络、多任务学习的不断深化,我们看到了使用一个框架来代替所有任务的曙光,它能将传统算法和现代学习融为一体,同时具有实时、泛化、宽基线以及一定的数学支持的特征,为真正的工业图像处理革命打下坚实的基础.

参考文献

References

[1] Fischler M A, Bolles R C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography[J]. Communications of the ACM, 1981, 24(6): 381-395

[2] Triggs B, McLauchlan P, Hartley R, et al. Bundle adjustment: a modern synthesis[C]// IEEE International Conference on Computer Vision, 1992: 98-372

[3] Moravec H. Obstacle avoidance and navigation in the real world by a seeing robot rover[R]. Tech Report, Robotics Institute, Carnegie Mellon University, 1980, CMU-RI-TR-01-18

[4] Harris C, Stephens M. A combined corner and edge detector[C]// Proceedings of the Alvey Vision Conference, 1988, DOI: 10. 5244/C.2. 23

[5] Shi J B, Tomasi C. Good features to track[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1994, DOI: 10. 1109/CVPR.1994. 323794

[6] Rosten E, Drummond T. Machine learning for high-speed corner detection[C]// European Conference on Computer Vision, 2006: 430-443

[7] 赵小川.现代数字图像处理技术提高及应用案例详解:MATLAB 版[M].北京:北京航空航天大学出版社,2012

[8] Marr D, Hildreth E. Theory of edge detection[J]. Proceedings of the Royal Society of London, Series B, Biological sciences, 1980, 207(1167): 187-217

[9] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110

[10] Lindeberg T. Scale-space theory in computer vision [M]. Berlin: Springer, 1994

[11] Bay H, Tuytelaars T, van Gool L. SURF: speeded up robust features[C]// European Conference on Computer Vision, 2006: 404-417

[12] Alcantarilla P F, Bartoli A, Davison A J. KAZE features [C]// European Conference on Computer Vision, 2012: 214-227

[13] Alcantarilla P F, Nuevo J, Bartoli A. Fast explicit diffusion for accelerated features in nonlinear scale spaces[J]. Proceedings of the British Machine Vision Conference, 2013. DOI: 10. 5244/C.27. 13

[14] Ke Y, Sukthankar R. PCA-SIFT: a more distinctive representation for local image descriptors[C]// Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004: 506-513

[15] 刘立,彭复员,赵坤,等.采用简化 SIFT 算法实现快速图像匹配[J].红外与激光工程,2008,37(1):181-184
LIU Li, PENG Fuyuan, ZHAO Kun, et al. Simplified SIFT algorithm for fast image matching[J]. Infrared and Laser Engineering, 2008, 37(1): 181-184

[16] Calonder M, Lepetit V, Strecha C, et al. Binary robust independent elementary features [C] // European Conference on Computer Vision, 2010: 778-792

[17] Rublee E, Rabaud V, Konolige K, et al. ORB: an efficient alternative to SIFT or SURF[C]// Proceedings of the 2011 International Conference on Computer Vision, 2011: 2564-2571

[18] Morel J M, Yu G S. ASIFT: a new framework for fully affine invariant image comparison[J]. SIAM Journal on Imaging Sciences, 2009, 2(2): 438-469

[19] 蔡国榕,李绍滋,吴云东,等.一种透视不变的图像匹配算法[J].自动化学报,2013,39(7):1053-1060
CAI Guorong, LI Shaozi, WU Yundong, et al. A perspective invariant image matching algorithm[J]. Acta Automatica Sinica, 2013, 39(7): 1053-1060

[20] Lin W Y, Wang F, Cheng M M, et al. CODE: coherence based decision boundaries for feature correspondence [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(1): 34-47

[21] Bian J W, Lin W Y, Matsushita Y, et al. GMS: grid-based motion statistics for fast, ultra-robust feature correspondence[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4181-4190

[22] Verdie Y, Yi K, Fua P, et al. Tilde: a temporally invariant learned detector [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 5279-5288

[23] Savinov N, Seki A, Ladicky L, et al. Quad-networks: unsupervised learning to rank for interest point detection[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1822-1830

[24] Barroso-Laguna A, Riba E, Ponsa D, et al. Key.Net: keypoint detection by handcrafted and learned CNN filters [J]. arXiv preprint, 2019, arXiv: 1904. 00889

[25] Zagoruyko S, Komodakis N. Learning to compare image patches via convolutional neural networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015: 4353-4361

[26] Simo-Serra E, Trulls E, Ferraz L, et al. Discriminative learning of deep convolutional feature point descriptors [C] // IEEE International Conference on Computer Vision, 2015: 118-126

[27] Han X F, Leung T, Jia Y Q, et al. MatchNet: unifying feature and metric learning for patch-based matching[C] // IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3279-3286

[28] Balntas V, Riba E, Ponsa D, et al. Learning local feature descriptors with triplets and shallow convolutional neural networks[C]// The British Machine Vision Conference, 2016. DOI: 10. 5244/C.30. 119

[29] Balntas V,Johns E,Tang L L,et al.PN-Net:conjoined triple deep network for learning local image descriptors[J].arXiv preprint,2016,arXiv:1601.05030
[30] Tian Y R,Fan B,Wu F C.L2-Net:deep learning of discriminative patch descriptor in euclidean space[C]// IEEE Conference on Computer Vision and Pattern Recognition,2017:661-669
[31] Mishchuk A,Mishkin D,Radenovic F,et al.Working hard to know your neighbor's margins:local descriptor learning loss[J].Advances in Neural Information Processing Systems,2017:4826-4837
[32] Luo Z,Shen T,Zhou L,et al.GeoDesc:learning local descriptors by integrating geometry constraints[C]// Proceedings of the European Conference on Computer Vision,2018:168-183
[33] Tian Y,Yu X,Fan B,et al.SOSNet:Second order similarity regularization for local descriptor learning[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2019:11016-11025
[34] Mishkin D,Radenović F,Matas J.Repeatability is not enough:learning affine regions via discriminability[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2018:287-304
[35] Ebel P,Mishchuk A,Yi K M,et al.Beyond cartesian representations for local descriptors[C]// Proceedings of IEEE International Conference on Computer Vision,2019:253-262
[36] Sivic J,Zisserman A.Video Google:a text retrieval approach to object matching in videos[C]// Proceedings of the Ninth IEEE International Conference on Computer Vision,2003:1470
[37] Moosmann F,Triggs B,Jurie F.Fast discriminative visual codebooks using randomized clustering forests[C]// Proceedings of the 19th International Conference on Neural Information Processing Systems,2006:985-992
[38] Sánchez J,Perronnin F,Mensink T,et al.Image classification with the fisher vector:theory and practice[J].International Journal of Computer Vision,2013,105(3):222-245
[39] Jégou H,Douze M,Schmid C,et al.Aggregating local descriptors into a compact image representation[C]// IEEE Conference on Computer Vision and Pattern Recognition,2010:3304-3311
[40] Arandjelovic R,Gronat P,Torii A,et al.NetVLAD:CNN architecture for weakly supervised place recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,2016:5297-5307
[41] Lin R C,Xiao J,Fan J P.NeXtVLAD:an efficient neural network to aggregate frame-level features for large-scale video classification[C]// Lecture Notes in Computer Science,2019:206-218
[42] Babenko A,Slesarev A,Chigorin A,et al.Neural codes for image retrieval[C]// Proceedings of the European Conference on Computer Vision,2014:584-599
[43] Babenko A,Lempitsky V.Aggregating deep convolutional features for image retrieval[C]// IEEE Conference on Computer Vision and Pattern Recognition,2016:4321-4329
[44] Gordo A,Almazán J,Revaud J,et al.End-to-end learning of deep visual representations for image retrieval[J].International Journal of Computer Vision,2017,124(2):237-254
[45] Husain S S,Bober M.REMAP:multi-layer entropy-guided pooling of dense CNN features for image retrieval[J].IEEE Transactions on Image Processing,2019,28(10):5201-5213
[46] Yi K M,Trulls E,Ono Y,et al.Learning to find good correspondences[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition,2018:2666-2674
[47] Sarlin P E,DeTone D,Malisiewicz T,et al.SuperGlue:learning feature matching with graph neural networks[J].arXiv preprint,2019,arXiv:1911.11763
[48] Yi K M,Trulls E,Lepetit V,et al.LIFT:learned invariant feature transform[C]// IEEE European Conference on Computer Vision,2016:467-483
[49] Ono Y,Trulls E,Fua P,et al.LF-Net:Learning local features from images[C]// Advances in Neural Information Processing Systems,2018:6234-6244
[50] Shen X L,Wang C,Li X,et al.RF-net:an end-to-end image matching network based on receptive field[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition,2019:8132-8140
[51] DeTone D,Malisiewicz T,Rabinovich A.SuperPoint:self-supervised interest point detection and description[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops,2018:337
[52] Revaud J,Weinzaepfel P,de Souza C,et al.R2D2:repeatable and reliable detector and descriptor[J].arXiv preprint,2019,arXiv:1906.06195
[53] Dusmanu M,Rocco I,Pajdla T,et al.D2-net:a trainable CNN for joint description and detection of local features[J].arXiv preprint,2019,arXiv:1905.03561
[54] Noh H,Araujo A,Sim J,et al.Large-scale image retrieval with attentive deep local features[C]// IEEE International Conference on Computer Vision,2017:3456-3465
[55] Teichmann M,Araujo A,Zhu M L,et al.Detect-to-retrieve:efficient regional aggregation for image search[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition,2019:5109-5118
[56] Luo Z X,Shen T W,Zhou L,et al.ContextDesc:local descriptor augmentation with cross-modality context[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition,2019:2527-2536
[57] Sarlin P E,Cadena C,Siegwart R,et al.From coarse to fine:robust hierarchical localization at large scale[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition,2019:12716-12725

南京信息工程大学学报(自然科学版),2020,12(3):261-273
Journal of Nanjing University of Information Science and Technology(Natural Science Edition),2020,12(3):261-273

A survey of image feature detection and matching methods

TANG Can[1] TANG Lianggui[1] LIU Bo[1]

1 School of Computer Science and Information Engineering, Chongqing Technology and Business University, Chongqing 400067

Abstract For decades, image feature detection and matching has been the foundation of computer vision. Without feature detection and matching, there would be no visual tasks such as SLAM, Sfm, AR, image retrieval, image registration, or panoramic images. Based on the review of classic detection algorithms in the past decades, this paper describes the latest progress in image feature detection and matching after the introduction of machine learning algorithm led by deep learning. The survey includes all the key points such as feature points, local descriptor, global descriptor, matching and optimization, and end-to-end framework, and compares the merits and demerits of each algorithm. In summary, facing the requirements of wide baseline, real-time, and low computing load detection from the industrial sector, image feature detection and matching is still a hard task. The multitasking global framework which fuses feature points, local descriptor, global descriptor, matching and optimization has become the trend of future research.

Key words image feature detection; descriptor; matching algorithm; deep learning