



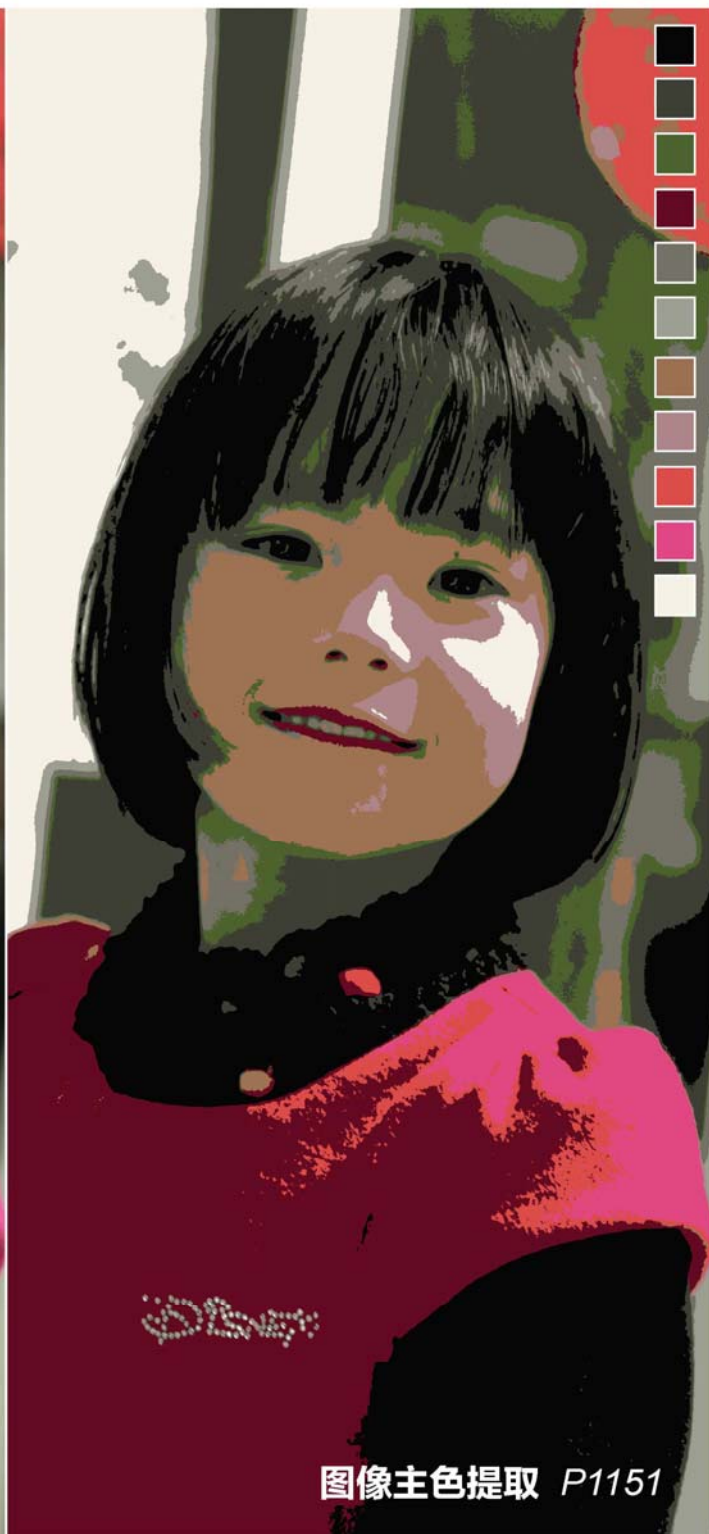
JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院遥感与数字地球研究所
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2015
09
VOL.20

ISSN1006-8961
CN11-3758/TB



图像主色提取 P1151



第20卷第9期（总第233期）

2015年9月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院

主办单位 中国科学院遥感与数字地球研究所

中国图象图形学学会

北京应用物理与计算数学研究所

主 编 李小文

编辑出版《中国图象图形学报》编辑出版委员会

邮政信箱 北京9718信箱

邮 编 100101

电子信箱 jig@radi.ac.cn

电 话 010-64807995

网 址 www.cjig.cn

广告经营许可证 京朝工商广字第0361号

总 发 行 北京报刊发行局

订 购 全国各地邮局

海外发行 中国国际图书贸易集团有限公司

（邮政信箱：北京399信箱 邮编：100048）

印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian

Monthly, Started in 1996

Superintended by Chinese Academy of Sciences

Sponsored by Institute of Remote Sensing and Digital Earth, CAS

China Society of Image and Graphics

Institute of Applied Physics and Computational Mathematics

Editor-in-Chief LI Xiaowen

Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics

P.O.Box 9718, Beijing, P.R.China

Zip code 100101

E-mail jig@radi.ac.cn

Telephone 010-64807995

Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals

Domestic All Local Post Offices in China

Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China))

Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB

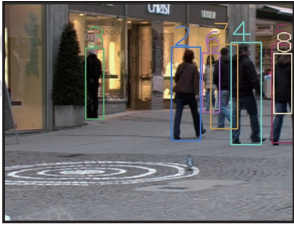
ISSN 1006-8961

CODEN ZTTXFZ

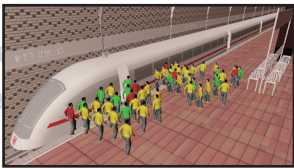
国外发行代号 M1406

国内邮发代号 82-831

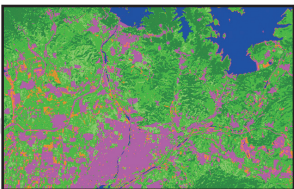
国内定价 50.00元



联合特征融合和判别性外观模型的多目标跟踪
(第1188页)



融入情绪模型的人群实时路径规划(第1262页)



基于Landsat 8卫星影像的北京地区土地覆盖分类
(第1275页)

综述

局部图像描述符最新研究进展

许允喜, 陈方 1133

图像处理和编码

主色提取的直方图峰值筛选与剔除方法

朱臻阳, 刘春晓, 伍敏, 陈丽丽 1151

强度和梯度稀疏约束下的图像平滑

胡大盟, 黄伟国, 张永萍, 杨剑宇, 朱忠奎 1161

基于稀疏编码的多模态信息交叉检索

刘菲, 刘学亮 1170

图像分析和识别

点线光流预测机制的图像序列运动直线跟踪

张聪炫, 陈震, 黎明 1177

联合特征融合和判别性外观模型的多目标跟踪

黄奇, 项俊, 侯建华, 张华, 笕邦友 1188

在线复合模板模型表示的视觉目标跟踪

亚森江·木沙, 赵春霞 1199

采用指数矩的图像区域复制粘贴篡改检测

赖玥聪, 黄添强, 蒋仁祥 1212

结合图元与感知哈希的手写输入简笔画识别

郭玉鹏, 曹卫群 1222

计算机图形学

网格修补的特征配准方法

钱归平, 张小红, 潘瑞芳 1230

基于生理舌头模型的语音可视化系统

江辰, 於俊, 罗常伟, 李睿, 汪增福 1237

面向移动终端的分布并行化渲染

刘镇, 刘晓, 梅向东 1247

反正切色调分化模型的铅笔画绘制

卢从利, 尚媛园, 丁辉, 周修庄, 付小雁 1253

虚拟现实与增强现实

融入情绪模型的人群实时路径规划

吴云鹏, 杜沅泽, 叶阳东 1262

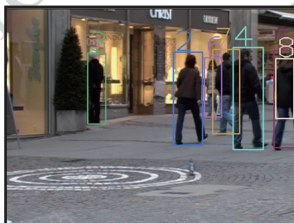
遥感图像处理

基于Landsat 8卫星影像的北京地区土地覆盖分类

王婷婷, 李山山, 李安, 冯旭祥, 吴业伟 1275

CONTENTS

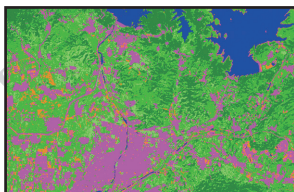
JOURNAL OF IMAGE AND GRAPHICS



Multi-target tracking algorithm based on feature fusion and discriminative appearance model (P1188)



Real-time emotion-integrated crowd path planning (P1262)



Land cover classification in Beijing using Landsat 8 image (P1275)

Review

- Recent advances in local image descriptor
Xu Yunxi, Chen Fang 1133

Image Processing and Coding

- Histogram peaks-filtering and rejection-based dominant color extraction
Zhu Zhenyang, Liu Chunxiao, Wu Min, Chen Lili 1151
- Image smoothing using intensity and gradient sparsity
Hu Dameng, Huang Weiguo, Zhang Yongping, Yang Jianyu, Zhu Zhongkui 1161
- Novel multi-modality information cross-retrieval based on sparse coding
Liu Fei, Liu Xueliang 1170

Image Analysis and Recognition

- Image sequence moving straight-line tracking method based on the forecasting mechanism of a point and straight-line optical flow
Zhang Congxuan, Chen Zhen, Li Ming 1177
- Multi-target tracking algorithm-based on feature fusion and discriminative appearance model
Huang Qi, Xiang Jun, Hou Jianhua, Zhang Hua, Da Bangyou 1188
- Online integrative template based model representation for visual object tracking
Yasin Musa, Zhao Chunxia 1199
- Image region copy-move forgery detection based on Exponential-Fourier moments
Lai Yuecong, Huang Tianqiang, Jiang Renxiang 1212
- Handwritten sketch recognition based on sketch entity and perceptual hashing
Guo Yupeng, Cao Weiqun 1222

Computer Graphics

- Mesh completion with feature registration
Qian Guiping, Zhang Xiaohong, Pan Ruifang 1230
- Speech visualization system based on physiological tongue model
Jiang Chen, Yu Jun, Luo Changwei, Li Rui, Wang Zengfu 1237
- Distributed parallel rendering method for mobile terminals
Liu Zhen, Liu Xiao, Mei Xiangdong 1247
- Pencil drawing method based on arctangent hue differentiation model
Lu Congli, Shang Yuanyuan, Ding Hui, Zhou Xiuzhuang, Fu Xiaoyan 1253

Virtual Reality and Augmented Reality

- Real-time emotion-integrated crowd path planning
Wu Yunpeng, Du Yuanze, Ye Yangdong 1262

Remote Sensing Image Processing

- Land cover classification in Beijing using Landsat 8 image
Wang Tingting, Li Shanshan, Li An, Feng Xuxiang, Wu Yewei 1275

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2015)09-1133-18

论文引用格式: Xu Y X, Chen F. Recent advances in local image descriptor[J]. Journal of Image and Graphics, 2015, 20(9): 1133-1150. [许允喜, 陈方. 局部图像描述符最新研究进展[J]. 中国图象图形学报, 2015, 20(9): 1133-1150.] [DOI: 10.11834/jig.20150901]

局部图像描述符最新研究进展

许允喜, 陈方

湖州师范学院信息工程学院, 湖州 313000

摘要: **目的** 局部图像描述符广泛应用于许多图像理解和计算机视觉应用领域, 如图像分类、目标识别、图像检索、机器人导航、纹理分类等。SIFT 算法的提出标志着现代局部图像描述符研究的开始。主要对最近发展的现代局部图像描述符进行了综述。**方法** 首先, 介绍了 4 大类局部图像描述符: 局部特征空间分布描述符、局部特征空间关联描述符、基于机器学习的局部描述符、扩展局部描述符 (局部颜色描述符、局部 RGB-D 描述符、局部空时描述符)。对局部图像描述符进行了分析和分类, 并总结了局部图像描述符的不变性、计算复杂度、应用领域、评价方法和评价数据集。最后, 展望了局部图像描述符的未来研究方向。**结果** 近年来局部图像描述符研究取得了很大进展, 提出了很多优秀的描述符, 在辨别性、鲁棒性和实时性方面有了很大提高, 应用领域不断拓展。**结论** 局部图像描述符应用广泛, 是计算机视觉领域的重要基础研究。而目前, 局部图像描述符还存在许多问题, 还需进一步的深入研究。

关键词: 局部图像描述符; 局部不变特征; 图像匹配; 图像分类

Recent advances in local image descriptor

Xu Yunxi, Chen Fang

Institute of Information Engineering, Huzhou University, Huzhou, Zhejiang 313000, China

Abstract: **Objective** Local image descriptors are widely utilized in many image understanding and computer vision applications, such as image classification, object recognition, image retrieval, robot navigation, and texture classification. The development of the SIFT algorithm highlighted the beginning of modern local image descriptor research. Recently developed modern local image descriptors are surveyed in this study. **Method** Four types of local image descriptors, namely, spatial distribution descriptors of local features, spatial correlation descriptors of local features, local descriptors based on machine learning, and extended local descriptors (local color descriptors, local RGB-D descriptors, and local space-time descriptors), are presented. The local image descriptors are then analyzed and categorized. The performance of the local image descriptors is investigated in terms of invariance, computation complexity, application field, evaluation methods, and evaluation datasets. Finally, this paper concludes with a discussion of directions for future research of the local image descriptors. **Result** In recent years, the research of local image descriptors has made great progress. Many excellent descriptors are proposed, and their performances have greatly improved in the distinctiveness, robustness and real-time, and their application fields are continually expanded. **Conclusion** Local image descriptors are widely utilized as an important and fundamental research field of computer vision. However, many problems in the use of these descriptors persist. This condition indicates that further research on local image descriptors is required.

收稿日期: 2015-03-25; 修回日期: 2015-05-20

基金项目: 国家自然科学基金项目 (61370173); 湖州市重点科技创新团队 (2012KC04)

第一作者简介: 许允喜 (1978—), 男, 讲师, 2004 年于苏州大学获通信与信息系统专业硕士学位, 主要从事计算机视觉方面的研究。

E-mail: xuyunxi@hute.zj.cn

Supported by: National Natural Science Foundation of China (61370173)

Key words: local image descriptor; local invariant feature; image matching; image classification

0 引言

与全局特征方法相比,局部特征对图像存在复杂背景干扰和部分遮挡的处理能力更强,且通常不需要图像分割或提取轮廓等预处理。因此,它在计算机视觉领域有非常广泛的应用。局部图像描述符用于表示局部图像块的辨别性信息,其对一些光学变化(如模糊、光照、噪声、JPEG 压缩)和几何变化(如旋转、尺度、视角)应保持一定的不变性,所以也称为局部不变描述符。局部图像描述符广泛应用于图像理解和计算机视觉领域,如从运动恢复结构^[1-3]、图像分类^[4-6]、纹理分类^[6-7]、目标识别^[8]、图像搜索^[9-11]、机器人导航^[12-14]、全景图拼接^[15]等。

在计算机视觉研究领域利用局部图像描述符主要有 3 种途径:1)传统方式,包含以下 3 个步骤:兴趣点检测、局部图像描述符提取和描述符匹配。主要用于特征跟踪或匹配、图像匹配、目标识别等视觉应用场合。2)图像块方式,主要包含 2 个步骤:稠密规则局部区域采样和局部区域图像描述符计算。主要用于纹理分类、目标检测、人脸识别等视觉应用。3)视觉单词包方式,包含以下 4 个步骤:特征采样(兴趣点检测、稠密规则图像块采样或随机图像块采样)、局部图像描述符提取、局部图像描述符聚类 and 局部图像描述符直方图构建。主要用于基于目标的图像搜索、图像分类等视觉应用。以前的局部图像描述符一般归类于局部不变特征(兴趣点检测和描述)。从局部图像描述符第 2 类和第 3 类应用情况看,局部图像描述符可以不需要局部兴趣点检测步骤,所以已经脱离局部不变特征研究,成为一个独立的研究课题。

一个理想的局部图像描述符应该具备高度的独特性、不变性和紧凑性。独特性是指:每个不同场景的局部图像描述符应具有很好的独特性,即能很好地辨别不同场景。不变性是指:局部图像描述符对同一场景不同光度和几何变换下应具有很好的不变性能。紧凑性是指:局部图像描述符的维数应远远小于图像块的像素数目,即局部图像描述符应具有低维性。

早期的局部图像描述符研究可以追溯到 80 年

代,而 SIFT 特征描述符^[8]的提出标志着局部图像描述符研究进入了一个全新的时期。SIFT(scale invariant feature transform)描述符提出后不久,Mikolajczyk^[16]对各类局部图像描述符(形状上下文^[17]、矩、复杂滤波器^[18]、互相关描述符、微分描述符、方向可调描述符^[19]、Spin 描述符^[20]、SIFT、PCA-SIFT(principal components analysis SIFT)^[21]、GLOH(gradient location and orientation histogram),其中,PCA-SIFT 和 GLOH 是 SIFT 的变体)进行了性能评价。从该评价结果看,SIFT 描述符及其变体在各种图像变换下的性能都优于其他类型的描述符。

SIFT 描述符的提出是局部图像描述符研究的一个里程碑,其标志着现代局部图像描述符研究的开始。近年来,局部图像描述符研究非常活跃,在计算机视觉主流国际会议(ICCV(IEEE International Conference on Computer Vision)、CVPR(IEEE Conference on Computer Vision and Pattern Recognition)、ECCV(European Conference on Computer Vision))以及一些国际权威期刊(如 IEEE Transactions on Pattern Analysis and Machine Intelligent, International Journal of Computer Vision, IEEE Transactions on Image Processing)上每年都有很多论文发表。本文主要针对这几年的最新研究成果进行综述。

1 局部图像描述符

1.1 局部特征空间分布描述符

局部特征空间分布描述符是目前应用最广泛的局部图像描述符,其提取像素层局部特征,然后通过空间分割方法对局部特征进行汇聚,得到像素层局部特征的空间分布。应用最广泛的局部特征空间分布描述符为基于梯度的描述符。SIFT 描述符是^[8]目前最流行的局部图像描述符,其是梯度位置和方向的 3D 直方图。SIFT 采用了 4×4 笛卡儿坐标网格的空间汇聚方法。SIFT 已成功用于目标检索、图像分类、机器人导航、图像致密对应^[22]等领域。针对 SIFT 描述符的高维度,Ke 和 Sukthankar^[21]提出了利用主分量分析(PCA)对 SIFT 进行维数约简得到更加紧凑的 PCA-SIFT 描述符。针对 SIFT 描述符计算复杂度高的问题,Bay 等人^[23]提出了 SURF

(speeded up robust features) 描述符, 利用 Box 滤波代替高斯滤波, 利用哈尔滤波器代替梯度。相对于 SIFT 描述符, SURF 描述符的计算速度大大提高。Mikolajczyk 等人^[16] 提出了 GLOH 描述符。GLOH 描述符是 SIFT 描述符的扩展, 鲁棒性和辨别性都有了提升。GLOH 描述符采用 3 个径向方向和 8 个角度方向的 log-polar 网格坐标。Engin 等人^[24] 提出了应用于宽基线立体和稠密特征提取的 DAISY 描述符。DAISY 描述符也计算梯度直方图, 但采用了新的空间汇聚方法。Dalal 等人^[25] 提出了方向梯度直方图(HOG)描述子并用于行人检测。HOG 描述子提取过程如下: 把样本图像分割为若干个像素的子区域, 把梯度方向平均划分为 9 个区间, 在每个子区域里面对所有像素的梯度方向在各个方向区间进行直方图统计。Takacs 等人^[26] 提出了旋转不变快速特征(RIFF)并用于实时跟踪和识别。RIFF 采用了旋转不变梯度计算和旋转不变空间汇聚策略。旋转不变梯度通过局部径向坐标系统计算得到。旋转不变空间汇聚策略通过 4 个圆环空间分割实现。RIFF 在不需要计算图像块的主方向情况下实现了旋转不变。Ambai 等人^[27] 提出了紧凑实时描述符(CARD)。CARD 描述符由简短的二进制编码组成, 可以非常快速的计算。提出了一种用于提取方向梯度直方图的基于查表的高效算法, 每个描述符的计算速度大约是 SIFT 的 16 倍。另外, 他们还提出了基于学习的稀疏索引方法把提取的描述符转化为二进制编码。

像素层局部特征有 1 维或多维, 如梯度特征包含 x 方向梯度和 y 方向梯度 2 个维度。局部描述符对这些 1 维或多维像素层特征一般都有一个编码量化的过程。Chandrasekhar 等人^[28] 对 (dx, dy) 的联合分布进行了统计分析, 并依据该分布特性对 (dx, dy) 2 维空间进行矢量量化。Boix 等人^[29] 提出了一种新的、更一般的基于稀疏量化的图像块描述形式, SIFT 的软分配编码和后面提到的 BRIEF 描述符的强度对比硬编码都可以看做它的实例。

Huang 等人^[30] 提出了对比度上下文直方图(CCH), 该描述符计算 32 个空间划分子区域的正负平均对比强度, 产生 64 维直方图向量。该描述符采用归一化的灰度特征。和采用梯度特征的 SIFT 描述符相比, CCH 计算复杂度和维数都较低。但 CCH 描述符的匹配性能和鲁棒性大大低于 SIFT 描

述符。

局部二进制模式是另一种像素层局部特征, 其编码中心像素和周围像素之间的相对强度值。Ojala 等人^[31] 提出了用于旋转不变纹理分类的基本局部二进制模式(LBP)方法。局部二进制模式描述符在纹理分类、人脸识别、图像匹配等领域表现出非常优越的性能。近年来局部二进制模式描述符的研究十分活跃, 出现了大量研究成果。在对比实验中, LBP 在计算效率、辨别性等方面都达到了很高的性能, 已成功应用于人脸识别^[32]、背景建模^[33]、表情识别^[34] 中。Wu 等人^[35] 提出的 CENTRIST 描述符中 Census 变换和 $LBP_{8,1}$ 是等同的。

自 Ojala 提出基本 LBP 算法以来, 有研究者提出了许多 LBP 改进形式用于人脸识别、纹理分类、特征匹配和目标识别中。为了提高 LBP 描述符的辨别性能, Guo 等人^[36] 提出了完全 LBP 算法。Tan 和 Triggs 把 LBP 扩展为三值编码, 即 LTP(local ternary pattern)^[37]。Zhang 等人^[38] 提出了人脸识别的局部微分模式 LDP(local derivative pattern)。LBP 是无方向的一阶局部模式操作符, LDP 把其扩展为高阶多方向。Rivera 等人^[39] 提出了局部方向数模式用于人脸和表情识别; Liao 等人^[40] 提出了关键局部二进制模式用于纹理分类; Zhao 等人^[41] 提出了完全局部二进制总数用于旋转不变纹理分类; Murala 等人^[42] 提出了局部四模式用于基于内容的图像检索; Maani 等人^[43] 提出了旋转不变局部频率描述符(LFD)用于纹理分类; Zhao 等人^[44] 提出了 LBP 直方图频率特征(LBP-HF)用于纹理分类; Ojansivu 等人提出了局部相位量化描述符(LPQ)并用于模糊纹理分类和模糊人脸识别^[45-47]。

上面介绍的局部二进制模式采用稠密规则采样图像块的图像区域描述方式, 主要用于纹理分类和人脸识别。为了使局部二进制模式用于稀疏特征匹配, Heikkila 等人^[48] 提出了中心对称局部二进制模式(CS-LBP)描述符, 在特征匹配和目标分类中获得了良好的性能。和 SIFT 描述符相似, CS-LBP 描述符在特征点周围取 4×4 小区域, 对这些小区域进行空间汇聚。和 CS-LBP 类似, Gupta 等人^[49] 在 LTP 基础上提出了 CS-LTP(center-symmetric local ternary pattern)描述符, 在大多数情况下其匹配性能都优于 CS-LBP 描述符。曾慧等人^[50] 提出了 ICS-LTP(Improved center symmetric local trinary pattern)算子, 采

用了加权纹理谱直方图计算方法,从而得到了 IWCS-LTP (improved weighted center symmetric local trinary pattern) 描述符。

Chen 等人^[51]提出了一种简单但功能强大的、鲁棒的韦伯局部描述符(WLD)。韦伯局部描述符来源于韦伯生物学准则,用当前像素和其邻居点之间的强度变化描述当前像素的变化,通过这种方法发现图像中的兴趣变化,从而仿真人类的模式感知。韦伯局部描述符包含二部分:差分激励(ξ)和方向(θ)。差分激励 ξ 即为当前像素点强度与其邻居点的强度差异的比率。方向分量 θ 即为局部梯度方向。在局部区域中对差分激励 ξ 和方向 θ 进行区间划分进行2维直方图统计即得到韦伯局部描述符。韦伯局部描述符应用广泛,除了用于人脸检测和纹理分类^[51]外在人脸识别^[52-53]、种族识别^[54]和性别识别^[55]等方面都得到了很好的应用。

近年来,针对梯度特征对非线性光照敏感的问题,研究者提出了利用强度序,即像素强度大小的排序信息来构建局部描述符。利用强度大小的排序信息可以编码每个像素在整个图像块或周围局部像素中的属性,其对单调光照变化具有不变性。许多复杂光照变化,如仿射、平方、平方根和其他非线性光照变化都属于这种类型。局部特征空间分布描述符一般利用强度序对图像空间进行分割。Gupta 和 Mittal^[56-57]把强度序信息用于特征匹配以提高匹配的鲁棒性。Tang 等人^[58]提出了强度序-空间分布描述符(ordinal spatial intensity distribution, OSID)。在序和图像空间上对像素强度进行分组,网格化2D-直方图即得到 OSID 描述符。OSID 描述符采用把归一化支撑圆形区域分割为若干扇形的图像空间分割方法。Fan 等人^[59]提出了基于序的多支撑区域梯度直方图描述符(MROGH)和基于多支撑区域的旋转和强度单调不变描述符(MRRID)。MROGH 描述符中基于梯度的局部特征计算和 SIFT 相似,但通过局部旋转不变坐标系统计算得到,其和 RIFF 描述符的局部径向坐标系统是等价的。MRRID 描述符中基于强度的局部特征和 CS-LBP 相似,但通过局部旋转不变坐标系统计算得到。在局部旋转不变坐标系统下基于梯度的局部特征和基于强度的局部特征都是旋转不变的。接着,旋转不变特征基于他们的强度序进行空间汇聚。因为这种汇聚策略也是旋转不变的,所以构建的描述符是旋转不变的,从

而避免了像 SIFT、CS-LBP、Daisy 等局部描述符估计图像块主方向的问题。MRRID 和 MROGH 描述符在不需计算图像块的主方向情况下实现了旋转不变,所以其辨别性很高。Wang 等人^[60]提出了局部强度序模式(LIOP)描述符。LIOP 采用 Fan 提出的基于图像块强度序的区域分割方法,提出了一种新的局部特征:局部强度序模式,并且在局部旋转不变坐标系统下进行计算。LIOP 描述符的局部强度序模式计算如下:在兴趣点周围半径为 R 的圆上等间隔取 N 个点,对 N 个点的强度进行排序,得到 $N!$ 种排序结果。Kim 等人^[61]提出了一种基于精确序的局部描述符(EOD)。EOD 对两种描述符 CS-LBP 和 OSID 进行改进和融合。

1.2 局部特征空间关联描述符

和局部特征空间分布描述符不一样,局部特征空间关联描述符不对支撑域进行空间分割,而是通过某种方法计算像素层局部特征(如梯度、颜色、局部二进制模式)的空间关联性。目前,局部特征空间关联描述符主要包含独立二进制强度对比描述符、核描述符、区域协方差描述符和基于矩的局部描述符4类。

1.2.1 独立二进制强度对比描述符

独立二进制强度对比描述符采用多个相互独立的2个像素间二进制强度对比来描述图像块,组成二进制比特串。由于其计算简单快速、存储容量低,匹配速度快,可以用于计算资源非常有限的移动设备上,近年来受到很多计算机视觉领域研究者的关注。

Calonder 等人^[62-63]提出了二进制鲁棒独立描述符(BRIEF)。BRIEF 描述符在局部图像块中采样若干个位置对,计算它们的二进制强度对比,采样位置对表达强度对比的空间关联。Galvezlopez 等人^[64]提出了利用 BRIEF 描述符的移动机器人位置识别方法。针对 BRIEF 描述符不具备旋转不变性及其强度对位置采样策略,有研究者提出了一些改进的独立二进制强度对比描述符。Rublee 等人^[65]提出了基于 BRIEF 的 ORB (oriented fast and rotated BRIEF) 二进制描述符。该描述符实现了旋转不变性,并提出了一种贪婪搜索的学习方法选择良好的位置采样。Stefan 等人^[66]提出了基于 BRIEF 的 BRISK (binary robust invariant scalable keypoints) 二进制强度对比描述符。该描述符实现了尺度不变和

旋转不变性,并提出了一种特别的关键点附近采样模式,从而最大化描述符的辨别性。Heinly 等人^[67]对 BRIEF、ORB、BRISK 3 种独立二进制强度对比描述符进行了性能评估:在大的旋转和几何变化条件下,BRIEF 的性能要低于 ORB 和 BRISK 描述符。受人类视网膜图像感知原理启发,Alahi 等人^[68]提出额 FREAK (fast retina keypoint) 描述符。该描述符采用视网膜形式的二进制位置对采样模式,视网膜采样格越接近中心其密度越高。利用高斯滤波器对每个采样点进行平滑以减少噪声的影响。每个采样点使用不同的高斯核,其大小根据视网膜模式呈极坐标增长。Ziegler 等人^[69]对 BRIEF 描述符及其改进描述符在新的视角上给出了分析和解释:他们都是在序相关性上的 Hash 方法。并提出了一种新的局部图像描述符:局部一致对比图像描述符(LUCID)。LUCID 描述符基于 2 个图像块 RGB 序之间的线性置换距离。

1.2.2 核描述符

Bo 等人^[70]指出:从核方法的视角看,方向直方图(如 SIFT、HOG)与图像块之间的某些匹配核是等价的。并从这种崭新的视角提出了一种统一的、规范的核描述符设计框架。该设计方法可以把像素层局部特征(梯度、颜色、局部二进制模式等)转化为紧凑的图像块层特征。提出了 3 类匹配核来度量图像块之间的相似性,并提出了一种在图像块中提取底层特征的统一框架:核描述符。

梯度匹配核 K_{grad} 定义为

$$K_{\text{grad}}(P, Q) = \sum_{z \in P} \sum_{z' \in Q} \tilde{m}(z) \tilde{m}(z') k_o(\tilde{\theta}(z), \tilde{\theta}(z')) k_p(z, z') \quad (1)$$

式中, P 和 Q 为来自两个不同图像的图像块, z 表示图像块中的像素位置。 $\tilde{\theta}(z)$ 和 $\tilde{m}(z)$ 是图像梯度在像素 z 处的归一化方向和幅度。 $k_p(z, z')$ 为高斯位置核。 $k_o(\tilde{\theta}(z), \tilde{\theta}(z'))$ 为高斯方向核。梯度特征的空间关联通过高斯位置核实现。上述的匹配核计算量非常大,Bo 等人采用了核主分量分析方法把匹配核转化为紧凑的低维核描述符。

核主分量分析方法为无监督维数约简学习方法,Wang 等人^[71]又提出了一种监督核描述符(SKDES)。监督核描述符把图像层标签信息嵌入到图像块层核描述符的设计中,采用大边界最近邻准则

的监督模型学习得到低维的核描述符。核描述符在目标识别、场景标注^[72]和视频分类^[73]等领域都获得了目前最好的结果。

1.2.3 区域协方差描述符

Tuzel 等人^[74]提出了区域协方差描述符,用于目标跟踪检测和图像纹理分类中^[75-77],获得了较好的效果。区域协方差实现了对图像多种底层特征的融合,对非刚体目标的几何变形、遮挡、视角以及光照变化等都有很强的适应性。

区域协方差描述符计算局部目标区域的 $d \times d$ 维协方差矩阵,该矩阵是实对称正定矩阵,矩阵中对角线上的值表示每个独立的特征量,非对角线上的值表示特征之间的相关量。 d 维特征为各种图像底层特征,如灰度、颜色、一阶梯度,二阶梯度等。为了编码底层特征在局部图像中的空间位置,底层特征通常还包括像素的图像空间位置,这样像素层局部特征的空间位置就通过协方差矩阵进行了关联。两个协方差矩阵之间的相似性度量不属于欧氏空间的问题,其采用特别的距离计算方法,该距离度量基于两个协方差矩阵之间的广义特征值。

1.2.4 基于矩的局部描述符

基于矩的局部描述符^[78-79]主要包含几何矩和复杂矩。由这些几何矩或复杂矩,可以构建整个图像块支撑域的旋转或仿射矩不变量。几何矩不变量通常仅适用于描述简单图像,而复杂矩不变量的辨别性能较弱。所以,早期基于矩的局部图像描述符在性能上大大低于主流的局部图像描述符,如 SIFT,其在计算机视觉领域的应用越来越少。

最近,Chen 等人^[80]提出了一种新的基于 Zernike 矩相位的局部图像描述符。该描述符对光度和几何变化很鲁棒,与 SIFT、PCA-SIFT、GLOH 相比,在各种光度和几何变化下,Zernike 矩相位描述符都获得了领先的性能。Zernike 矩^[81]可以看成几何矩的扩展,即把传统几何矩的变换核 $x^m y^n$ 替换为正交 Zernike 多项式。阶数为 n 和副本为 m 的 Zernike 基函数 $V_{nm}(\rho, \theta)$ 的定义为^[81]

$$V_{nm}(\rho, \theta) = R_{nm}(\rho) e^{im\theta}, \rho \leq 1 \quad (2)$$

式中, $R_{nm}(\rho)$ 为径向多项式, ρ 为极坐标下极径, θ 为极角。对局部图像块进行 Zernike 矩计算, n 和 m 取不同的值时可以得到不同的 Zernike 矩。Zernike 矩相位局部图像描述符只计算 Zernike 矩的相位角。

1.3 基于机器学习的局部描述符

手动设计的描述符很难确定参数配置,设计出的描述符很难同时具备很高的独特性、不变性和紧凑性。近几年,研究者把机器学习方法用于描述符设计并成为局部不变描述符研究领域的一大热点。相对于手动设计,基于机器学习的局部不变描述符设计方法有其独特的优势:可以利用训练数据学习最佳的参数配置和复杂模式。本文把基于机器学习的局部不变描述符分为基于机器学习的浮点数局部描述符和基于机器学习的二进制局部描述符。

Brown 等人^[82-83]提出了一种统一的局部图像描述符构建框架:局部图像描述符可以由若干模块组合得到。并提出了构建局部不变描述符的学习框架:把构建描述符的模块组合在一起,采用匹配/非匹配图像块训练数据和机器学习算法通过使匹配/非匹配分类性能最佳来优化这些模块的参数。

Brown 提出的局部描述符学习包括 4 个构建模块:G-Block:对输入图像块进行高斯平滑操作;T-Block:对平滑图像块进行非线性变换,主要包含各种图像滤波器;S-Block/E-Block:对上述滤波器响应进行空间汇聚操作。S-Block 使用参数化的汇聚区域,E-Block 为非参数化的;N-Block:对描述符进行归一化去除光照影响。Brown 联合这些模块建立目标优化函数,采用图像块训练数据集和 Powell 优化技术最小化最近邻分类器得到最佳的参数配置。

Brown 采用的优化准则不是解析的,对应的优化目标函数为非凹的,容易陷入局部最小。针对这些缺点,Simonyan 等人^[84]提出了基于凸优化的局部描述符学习方法:1)把描述符的空间汇聚学习描述为一个凸优化问题并可以采用稀疏性选择汇聚区域;2)描述符维数约简也可以描述为一个凸优化问题,并采用核范数约简维数。Trzcinski 等人^[85]提出了一种低维增强梯度映射(LBGM)的局部描述符学习方法,其提供了一种通用的学习框架,而且局部图像描述符优化形式更简单。该方法包含 3 部分:1)基于梯度的弱学习器;2)适合于低维、辨别嵌入增强学习的相似度函数;3)对嵌入进行参数化形成紧凑的描述符(相当于维数约简)。

针对辨别性维数约简方法,Mikolajczyk 等人^[86]和 Hua 等人^[87]提出了两种新的辨别嵌入方法用于描述符维数约简。Cai 等人^[88]证明了二种辨别嵌入方法是等价的,并统一称为线性辨别投影(LDP)。

LDP 使不同标签点方差和相同标签点方差比率最大化。徐小明等人^[89]提出了基于局部不变映射(LPP)的局部描述符设计算法。该算法用 LPP 预先生成一个特征矩阵,接着把支撑域内所有点的梯度组成一个高维的梯度向量,最后通过特征矩阵把该梯度向量嵌入到一个低维的流形空间中。该算法不仅考虑了特征点的邻接点信息,而且还保持了特征点的内在邻接属性不变和特征点的可区分性,所以在保持描述符低维数的同时也获得了很高的性能。

在大尺度检索和匹配问题中,像高维的浮点数 SIFT 描述符存在存储容量大和匹配效率低的问题。前面介绍的独立二进制强度对比描述符存储容量低、匹配效率高,但其辨别性不高。近年来,研究者开始利用机器学习方法得到二进制比特串描述符。学习得到的二进制比特串描述符不仅存储容量低、匹配效率高,辨别性能也很高。Strecha 等人^[90]提出了利用机器学习方法把 SIFT 描述符投影量化为二进制比特串描述符,称为 LDAHash (linear discriminant analysis hash) 描述符。Trzcinski 等人^[91]提出了基于图像块的二进制描述符学习方法,称为 D-Brief (discriminative BRIEF) 描述符。D-Brief 描述符与 LDAHash 描述符不同,LDAHash 在 SIFT 描述符之上进行投影量化的机器学习,限制了 LDAHash 的精度,而 D-Brief 描述符直接在图像块上进行投影量化的机器学习,所以辨别性更高。Trzcinski 等人^[92]引入了基于 Boost 的二进制描述符学习,Boost 方法同时学习每个比特的特征最佳权重和形状。实验结果表明:8 比特就能得到 SIFT 相似的性能,增加比特长度能大大提高性能。

1.4 扩展局部描述符

1.4.1 局部颜色描述符

上述描述符都建立在灰度图像上,而目前绝大多数图像传感器都是彩色的。人类会充分利用颜色信息进行图像理解,颜色信息将有助于目标识别、图像分割、图像匹配等。近年来,局部描述符研究已扩展到颜色图像。目前,局部颜色描述符主要可以分为以下 3 类:

1)颜色描述符直接统计各种颜色通道形成颜色直方图,主要包括 RGB 直方图,对立直方图,色调直方图,归一化 rg 直方图和变换的颜色直方图。

2)局部颜色描述符将各个颜色通道分别用于形状(灰度)描述符,主要包括 RGB-SIFT、rgSIFT、

HSV-SIFT^[93]、OpponentSIFT^[94]、C-SIFT^[95]等。其中, HSV-SIFT 是 H/S/V 三个通道分别计算 SIFT 描述符。OpponentSIFT 分别对对立颜色空间中的 O_1 , O_2 , O_3 计算 SIFT 描述符。C-SIFT 消除了 O_1 和 O_2 中的强度信息, 采用了归一化的对立颜色空间 O_1/O_3 和 O_2/O_3 。另外, Weijer 等人^[96]提出了对光度变化鲁棒的局部颜色张量。Brown 等人^[97]提出了一种多光谱 SIFT 用于场景分类识别, 该描述符将一个 4D 的 RGB + NIR 颜色向量和 SIFT 描述符合并用于表达空间和颜色信息。

3) 颜色描述符直接把颜色直方图和灰度描述符进行级联, 如 HueSIFT^[98] 将灰度级的 SIFT 描述符和色调直方图级联。

文献^[94]对各种颜色描述符的不变性和辨别性进行了评价, 总结了在各种线性光照变化下的不变性。图像和视频基准数据库的分类识别对比实验表明: 在 SIFT 基础上加入颜色描述符, 识别率会大大提高; 在缺乏数据库先验知识情况下, OpponentSIFT 是最好的选择。最近, Zhang 等人^[99]提出了一种新的空间—颜色对立编码理论: 用于颜色表面的 SO 编码和用于颜色边缘的 DO 编码, 建立了 SO (single-opponent) 和 DO (double-opponent) 层次连接的编码方式。这种层次处理优于颜色和形状 (灰度) 信息的简单级联, 并得到 SOSIFT、DOSIFT 以及 SODOSIFT 局部颜色描述符。

1.4.2 局部 RGB-D 描述符

在计算视觉领域, 局部描述符另一个研究热点为 3D 局部描述符。3D 局部描述符依据其是否使用局部坐标帧分为两类。不使用任何局部坐标帧的描述符采用局部几何统计或直方图形成特征描述符。不使用任何局部坐标帧的描述符在形成直方图过程中丢失了大多数 3D 空间信息, 所以其辨别性受到限制。使用局部坐标帧的特征描述符根据定义的局部坐标帧编码邻居点的空间分布和几何信息。然而没有一个局部坐标帧定义能同时满足唯一性、对噪声和网格分辨率鲁棒性的条件, 这大大影响了使用局部坐标帧的特征描述符的鲁棒性和辨别性。

随着便宜、实时的 RGB-D 传感器 (如微软的 Kinect) 的广泛应用, 获得高质量的颜色和深度同步数据变得很容易, 所以近年来联合几何信息和颜色 (灰度) 信息的 RGB-D 局部描述符成为智能机器人

和计算机视觉领域的研究热点。Blum 等人^[100]提出了卷积 k-均值局部 RGB-D 描述符并用于目标识别和位姿估计中。Bo 等人^[101-102]提出了 RGB-D 深度核描述符。Tombari 等人^[103]提出了 CSHOT (color-SHOT) 局部描述符, 该描述符是 SHOT (signature of histograms of orientations)^[104]的扩展。CSHOT 描述符有两个直方图构成: 球体支撑区域上的几何特征直方图、每个邻居点颜色空间的绝对差值之和的纹理特征直方图。CSHOT 描述符直接由几何特征直方图和纹理特征直方图级联而成。Nascimento 等人^[105]提出了融合独立二进制强度对比描述符和几何信息的 BRAND (binary robust appearance and normals descriptor) 局部描述符, 该描述符获得了非常好的匹配效果在评价实验中, 其匹配性能大大超过了 SURF 描述符和 SIFT 描述符。BRAND 局部描述符由 256 位二进制比特串组成。

1.4.3 局部空时描述符

最近, 局部图像描述符已成功扩展到空时领域, 局部空时描述符已成为行为识别和视频分析中流行的视频表示方法^[106-108]。目前, 局部空时描述符的研究主要分为以下 3 类:

1) 局部空时描述符很简单, 其对 3D 空时块的空间和时间属性进行独立计算, 然后进行合并得到描述符。Laptev 等人^[109]提出的 HOG/HOF 就属于这类空时描述符。

2) 空时描述符把局部描述符的计算过程从 2D 域扩展到 3D 域。这类描述符的主要思想为: 用 3D 空时块代替 2D 图像块, 即把 (x, y) 空间域的描述符计算流程转化为空时域 (x, y, t) , 这类空时描述符研究的比较多, 主要包括 3D-SIFT^[110], HOG-3D^[111], VLBP^[112] 和 ESURF^[113]。3D-SIFT 是由 SIFT 描述符扩展得到, HOG-3D 由 HOG 描述符扩展得到, VLBP 由 LBP 描述符扩展得到, ESURF 由 SURF 描述符扩展得到。

3) 局部空时描述符在 3D 空时块的 3 个正交平面上提取一组 2D 局部描述符, 称为 TOP (three orthogonal planes) 空时描述符。这类空时描述符主要包括 LBP-TOP^[114]、WLD-TOP^[115]、LQP-TOP^[116]。LBP-TOP 由 LBP 描述符扩展得到, WLD-TOP 由 WLD 描述符扩展得到, LQP-TOP 由 LQP 描述符扩展得到。Norouznezhad 等人^[117]把 TOP 方法扩展到 9 个对称平面 (NSP), 除了 3 个正交平面外还利用

了 6 个对角平面以更好地建模视频块的动态信息,提出了 HOG-NSP 空时描述符。

最近,针对 RGB-D 传感器,提出了融合深度信息的 4D 局部空时描述符^[118-119]。

2 局部图像描述符分析、分类和总结

2.1 局部图像描述符分析和分类

把局部图像描述符分为 4 大类:局部特征空间分布描述符、局部特征空间关联描述符、基于机器学习的局部描述符、扩展局部描述符(局部颜色描述符、局部 RGB-D 描述符、局部空时描述符)。其中,局部特征空间分布描述符的研究最为活跃,描述符数量众多。下面从 4 个不同方面(像素层局部特征提取方法、局部特征量化方法、局部特征空间汇聚方法、描述符编码方法)对局部特征空间分布描述符进行分析和分类,表 1 给出了分类结果。

目前,最常用的像素层局部特征有 5 种:1) 梯度,如 SIFT、DAISY、SURF;2) 归一化灰度,如 CHH;3) 局部二进制模式,如 CS-LBP;4) 局部韦伯描述符,如 WLD;5) 局部强度序,如 LIOP。

像素层局部特征一般都有一个特征空间分割过程,也称为特征量化。目前,局部图像描述符的量化方法主要有标量量化联合、软分配、矢量量化和稀疏量化 4 种。例如,局部二进制模式和局部韦伯描述符都采用标量量化联合,即像素层局部特征的每个维数分别进行标量量化,然后进行多维联合。SIFT 描述符采用软分配的方式对 2 维梯度进行量化。CHoG 描述符采用矢量量化方式对 2 维梯度进行量化。Boix 等人提出了稀疏量化方式。

像素层局部特征空间汇聚方法主要有图像空间手动分割、强度序空间分割。

图像空间手动分割主要包含 5 种方式:笛卡儿坐标的 2 维网格、log-polar 坐标的 2 维网格、Daisy 空间划分、扇形划分(OSID)、RIFF 圆环空间划分。目前,绝大多数描述符都是通过像素层局部特征在分割的空间上进行汇聚形成直方图。如流行的 SIFT、SURF、GLOH、Daisy、RIFF、CS-LBP 等描述符都采用图像空间手动分割方法。其中,RIFF 圆环空间划分是旋转不变的,其不需要计算图像块的主方向。其他空间手动分割方法需要计算图像块的主方向并把图像块绕主方向旋转来保持旋转不变。

强度序空间分割利用像素强度在整个图像块的序对图像空间进行分割,如 MROGH、MRRID 和 LIOP 描述符都采用该方法。强度序空间分割方法是旋转不变的,其不需要计算图像块的主方向。

表 1 局部图像描述符分类

Table 1 Classification of local image descriptors

方法	局布特征
像素层局部特征提取	梯度
	归一化灰度
	局部二进制模式
	局部韦伯描述符
局部特征量化	局部强度序
	标量量化联合
	软分配
	矢量量化
局部特征空间汇聚	稀疏量化
	笛卡儿坐标的 2 维网格
	log-polar 坐标的 2 维网格
	Daisy 空间划分
描述符编码	扇形划分
	RIFF 圆环空间划分
	强度序空间划分
	浮点数
描述符编码	维数约简
	二进制编码

描述符编码方法主要分为 3 种:浮点数、维数约简、二进制编码。第 1 种方法就是空间会聚得到的原始归一化浮点数描述符,如 SIFT。第 2 种方法就是对原始浮点数描述符利用维数约简方法对其维数进一步约简,如 PCA-SIFT。第 3 种方法就是把原始浮点数描述符编码为二进制描述符,如 CARD。

独立二进制强度对描述符计算一系列两个像素位置强度的对比,把两个像素强度的对比和两个位置都编码在二进制比特串中,实现了对像素层局部特征和像素位置关联的目的。核描述符和区域协方差描述符把像素层局部特征和空间位置利用核方法和协方差方法关联在描述符中,并采

用特定的相似性度量计算描述符之间的距离。基于矩的局部描述符把像素层局部特征和像素位置都编码在矩中。

目前的机器学习描述符都是基于局部特征空间分布描述符的研究框架,利用机器学习方法在像素层局部特征提取、局部特征量化、局部特征空间汇聚、描述符编码方法这4个层面进行优化设计。

计算描述符之间的相似性度量方法分为直方图对比方法、汉明距离方法、核方法、协方差距离。直方图对比方法又分为欧氏距离、直方图交叉核^[120]、Earth Mover 距离^[121]。汉明距离用于二进制比特串描述符的距离度量。核方法和协方差距离分别对应于核描述符和区域协方差描述符的相似性度量计算。

2.2 局部图像描述符的独特性和不变性分析

独特性(辨别性)和不变性是局部图像描述符设计的2个重要指标。独特性指局部图像描述符对不同外观的区域应具有独特性和区分性,即能很好的辨别不同的局部区域。局部图像描述符的不变性主要包括几何变化不变性和光度(光照)变化不变性。几何变化主要包括尺度变化、旋转变换、仿射变化、和视角变化(投影变换),这几种几何变化的变换矩阵为

$$T_s = \begin{bmatrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad T_r = \begin{bmatrix} \cos(\theta) & \sin(\theta) & 0 \\ -\sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$T_a = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \quad T_h = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix}$$

局部图像描述符的尺度不变\仿射不变一般由局部尺度\仿射不变特征提取和区域归一化完成,属于局部不变特征提取的研究范畴。描述符的旋转不变性一般通过以下两种方法实现:

1)描述符内在具有旋转不变性。几何矩和复杂矩描述符。局部特征空间分布描述符(RIFF、MR-RID、MROGH、LIOP)通过圆环或强度序这种旋转不变空间分割方法和局部旋转不变坐标系统计算像素层局部特征达到内在的旋转不变性。

2)通过计算局部图像块的参考方向然后图像块绕参考方向旋转的方法完成。参考方向估计一般采用梯度方向直方图方法,但该方法在某些情况下误差较大。

线性光度变化可统一表达为

$$\begin{bmatrix} R^c \\ G^c \\ B^c \end{bmatrix} = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} o_1 \\ o_2 \\ o_3 \end{bmatrix}$$

根据 a 、 b 、 c 和 o_1 、 o_2 、 o_3 的取值不同,把光度变化分为以下5种:

1)光照强度变化,即

$$a = b = c, o_1 = 0, o_2 = 0, o_3 = 0$$

2)光照强度平移,即

$$a = 0, b = 0, c = 0, o_1 = o_2 = o_3$$

3)光照强度变化和位移,即

$$a = b = c, o_1 = o_2 = o_3$$

4)光照颜色变化,即

$$o_1 = 0, o_2 = 0, o_3 = 0$$

5)光照颜色变化和位移。

表2给出了主流的局部颜色描述符的光度不变性分类。

表2 局部颜色描述符的光度不变性分类

Table 2 Photometric invariance classification of local color descriptors

	光照强度变化	光照强度平移	光照强度变化和位移	光照颜色变化	光照颜色变化和位移
SIFT(灰度图像)	Y	Y	Y		
HSV-SIFT					
RGB-SIFT	Y	Y	Y	Y	Y
C-SIFT	Y				
OpponentSIFT	Y	Y	Y		
rgSIFT	Y				

前3种线性光度变化对应于灰度图像的线性光照变化,即 $I_2 = kI_1 + b$ 。本文把灰度级2D图像局部描述符的光照不变分为线性光照不变和单调非线性光照不变两种。绝大多数描述符都具备线性光照不变性,一般通过像素相减(如梯度)和描述符归一化实现。像素相减去除了加性光照 b 的影响,归一化去除了乘性光照 k 的影响。假设图像块中任意2个像素强度满足 $I_1 > I_2$,经过非线性光照变换 $f(x)$ 后,如果满足 $f(I_1) > f(I_2)$,则满足单调非线性光照不变性。利用强度序和强度对比信息的描述符都满足单调非线性光照不变性。表3给出了主要灰度级2D图像局部描述符的线性光照不变性、单调非线性光照不变性和旋转不变性(指描述符内在的旋转不变性)以及视角变化鲁棒性、独特性的分类。表3中视角变化鲁棒性和独特性依据发表文献的评价结果分析总结得到:RIFF、CHH和独立二进制强度对比描述符的特点为维数低、计算速度快,但整体匹配性能中等;OSID和LIOP采用强度序全局光度属性,所以其视角变化鲁棒性中等;WLD和HOG主要用于目标检测,所以其独特性较低;区域协方差和核描述符对像素层多种特征进行融合,能用于变形目标的识别和检测,所以其视角变化鲁棒性很高,但独特性中等;根据文献[80],几何矩和复杂矩描述符视

角变化鲁棒性和独特性都较低;其他描述符都具有高的视角变化鲁棒性和独特性。2D灰度图像的局部描述符性能评价一般都采用文献[16]中给出的图像数据集和方法,根据变换图像和基准图像之间的局部描述符匹配来进行不变性和独特性评价。根据文献[16],视角变化仅指不包含摄像机平面内旋转的拍摄视角变化。

2.3 局部图像描述符的计算复杂度分类

把局部图像描述符的计算复杂度分为低、中、高和很高4种,表4给出了局部图像描述符计算复杂度的分类。

2.4 局部图像描述符的主要应用领域总结

把局部图像描述符的主要应用领域分为纹理分类、特征匹配、图像分类、目标识别和目标检测五类。表5给出了局部图像描述符的主要应用领域分类。特征匹配是局部图像描述符最主要的也是最基本的应用领域。特征匹配包含3个步骤:局部特征提取、局部描述符提取、描述符匹配。特征匹配主要包含以下应用于图像匹配、图像配准、图像拼接、从运动恢复结构、图像搜索等。特征匹配的研究热点主要针对大规模图像搜索中高效的描述符存储和匹配方法,如分层 k 均值聚类^[4]、汉明嵌入^[122]、乘积量化^[123]。

表3 局部图像描述符的不变性和独特性分类

Table 3 Invariance and distinctiveness classification of local image descriptors

	线性光照不变性	单调非线性光照不变性	旋转不变性	视角变化鲁棒性	独特性
SIFT、SURF	Y			高	高
WLD、HOG	Y			高	低
RIFF	Y		Y	中	中
CHH	Y			中	中
独立二进制强度对比描述符	Y	Y		中	中
CS-LBP	Y	Y		高	高
OSID	Y	Y		中	高
EOD	Y	Y		高	高
MRRID	Y	Y	Y	高	高
MROGH	Y		Y	高	高
LIOP	Y	Y	Y	中	高
几何矩和复杂矩描述符			Y	低	低
Zernike 矩相位描述符	Y			高	高
区域协方差、核描述符				很高	中
基于学习的描述符	Y			高	高

表4 局部图像描述符的计算复杂度分类

Table 4 Computation complexity classification of local image descriptors

	低	中	高	很高
独立二进制强度对比描述符	Y			
RIFF、CARD	Y			
CHH	Y			
SIFT			Y	
SURF		Y		
LBP、CS-LBP		Y		
WLD		Y		
OSID、EOD			Y	
MROGH、MRRID			Y	
区域协方差				Y
核描述符				Y
基于矩的描述符			Y	
基于机器学习的浮点数局部描述符			Y	
基于机器学习的二进制比特串描述符	Y			
局部 RGB-D 描述符			Y	
局部空时描述符				Y

表5 局部图像描述符的应用领域分类

Table 5 Application field classification of local image descriptors

	纹理分类	特征匹配	图像分类	目标识别	目标检测
WLD	Y				Y
SIFT、SURF		Y	Y	Y	
HOG					Y
LBP	Y		Y	Y	
CARD		Y			
独立二进制强度对比描述符		Y			
CS-LBP		Y		Y	
核描述符				Y	
区域协方差描述符				Y	Y
CHH、OSID、EOD		Y			
MROGH、MRRID		Y		Y	
Zernike 矩相位描述符		Y			
基于学习的描述符		Y		Y	

局部描述符应用于图像分类主要采用词包模型 (bag-of-words)^[124]。该模型的基本思想是将图像描述成一组无序局部描述符的集合:对一幅图像先提取若干局部描述符,然后将其映射到字典中的单词,该字典通常由聚类方法生成,然后将该图像表示为

单词出现频率的直方图。所以,词包模型一般由以下4部分组成:局部描述符提取、字典构造、基于字典的描述符编码、分类器设计。因为在图像分类中字典的单词数量通常非常庞大,基本的基于向量量化的描述符编码方法性能受限,所以产生了许多改

进的基于字典的描述符编码方法,如核词典编码^[125]、局部线性约束编码^[5]、Fisher 向量编码^[126]、局部聚合描述符向量编码^[127]、超向量编码^[128]。

纹理分类即对组成图像纹理的基元进行分类。在纹理分类领域中,词包模型方法也很流行^[6],从纹理图像上学习得到的词称为纹元(texton),所有纹元构成的字典称为纹元字典(texton dictionary)。纹元字典中的每一种纹元代表着一种纹理基元。纹理分类领域中各种字典学习方法的主要不同在于提取的局部描述符,不同的局部描述符会产生不同特性的字典。不像图像分类、纹理分类中字典的纹元数量通常不大。

目标检测任务与图像分类任务最重要的不同在于:目标结构信息在目标检测中起着至关重要的作用。目标检测的输入是包含目标的窗口,通过滑动窗口方法在输入图像的多个尺度上进行滑动扫描,而图像分类的输入则是整个图像。目标检测一般是指某一类目标的检测(如行人检测、人脸检测),以区分于目标识别。目标识别指对某一指定目标的识别(如人脸识别)。

2.5 局部图像描述符评价方法和评价数据集

评价局部图像描述符最常用的方法就是利用图像匹配进行性能评价。其中,牛津大学 Mikolajczyk 的图像匹配数据集和评价方法^[16]应用最为广泛。该数据集包含图像旋转、尺度变化、视角变化、图像模糊、光照变换和 JPEG 图像压缩六种图像变换。描述符性能评价采用 recall 对 1-precision 曲线。

Mikolajczyk 的评价数据集限于平面场景图像,没有复杂的视角变化,没有光源位置的变化。针对这些问题,Winder 等人^[82]提出采用图像块来评价描述符的匹配性能,这些数据集包含来自于不同视角和光照条件下的图像块,使用摄像机标定和多视立体得到基准数据。其采用正确匹配率和错误肯定率的 ROC(receiver operating characteristic)曲线评价局部描述符的性能。

Moreels 等人^[129]建立了在 144 个标定的视角和 3 个不同光照条件下的 100 个 3D 物体的图像数据库。其采用检测率对虚警率的 ROC 曲线来评价描述符的匹配性能。

为了得到精确可控的环境,Aanaes 等人^[130]建立了 DTU 数据集。该数据集包含 60 个不同的场景和不同类别的目标,每个场景有 119 个摄像机位置,每

个位置配有 19 个单独的 LED 进行光照控制。Kaneva 等人^[131]使用一个逼真的虚拟世界来得到不同光照条件和视角的可控环境来评价局部图像描述符。

上述都是单个图像之间的匹配,一些视觉应用是序列图像匹配问题。所以,针对基于特征的视觉跟踪,Gauglitz 等人^[132]提出了视频序列数据库。针对视觉同时定位与地图构建(vSLAM)应用,Gil 等人^[133]提出了导航序列图像数据库以及针对序列图像的局部图像描述符评价方法。

3 研究展望

利用计算机视觉技术来模仿人类感知能力的一个核心问题是:如何对所识别的物体进行表示。局部图像描述符的出现,为解决这一问题提供了有效手段。本文主要对最近发展的现代局部图像描述符进行了综述、分类和分析。尽管目前局部图像描述符的研究取得了很大的进步,但仍存在很多挑战和发展机遇。

1)人的视觉信息处理过程是分级的:从低级的 V1 区提取边缘特征,再到 V2 区的形状或者目标的部分等,再到更高层(如整个目标、目标的行为)。也就是说高层的特征是低层特征的组合,从低层到高层的特征表示越来越抽象,越来越能表现语义。而抽象层面越高就越利于分类。现有的局部图像描述符缺乏语义层次的解释,对场景或图像很难形成足够的描述力。提取包含更加丰富信息的中层特征(目标部件级别的局部图像块描述符)将是解决目前局部图像描述符缺乏语义信息的一个途径。所以,中层局部图像块的提取和描述是未来一个很重要的研究方向。

近年来流行的深度学习将是一个非常前途的解决方法。深度学习通过组合低层特征形成更加抽象的高层表示属性类别或特征,利用模拟人脑进行分析学习的神经网络发现数据的分布式特征表示。深度学习本身具有分层结构,其模拟大脑的分层思维,所以用深度学习方法训练底层的局部特征形成目标部件级的中层特征是完全可行的。深度学习的概念由 Hinton 等人于 2006 年提出,目前主要有深信度网、限制波尔兹曼机、多层自动编码器深层结构和卷积神经网络几种结构。其中,卷积神经网络是

第一个真正多层结构学习算法,它利用空间相对关系减少参数数目以提高训练性能,可能将是解决目前局部图像描述符缺乏高层语义信息这一问题最有前景的深度学习模型。

2)局部图像块描述符除了具备高辨别性外还应具备一定的不变性和紧致性(低维性)。这几种性质是互相关联的,手动设计很难达到平衡。例如,手动设计的高辨别性局部描述符的维数一般很高(如SIFT),存在一定的冗余信息和分量之间存在较强的相关性(紧致性差),手动设计的低维局部描述符辨别性一般较低。并且手动设计的局部描述符不能自适应于具体视觉应用。基于上述分析,结合具体视觉应用研究基于机器学习的局部图像块描述方法是将来的研究热点,因为通过机器学习方法可以建立合适的优化目标函数按照具体视觉应用要求同时优化局部描述符的辨别性、不变性和紧致性。目前,基于机器学习的局部图像描述符还缺乏系统深入的研究,其利用的像素层特征都是梯度信息,主要研究基于机器学习的空间汇聚策略。基于监督学习的维数约简和度量学习方法将是未来研究基于学习的局部描述符的一个重要研究方向,其可以使匹配的描述符靠拢(优化描述符的不变性),而同时使非匹配的描述符尽量远离(优化描述符的辨别性)以及降低描述符维数(优化描述符的紧凑性)。在基于监督式描述符学习方法研究中要构建大规模的基准图像块匹配数据库。手动标注这种大规模的基准匹配数据库的工作量是巨大的,所以采用多视几何等方法进行自动标注得到基准图像块匹配数据库也是基于学习的局部描述符的一个研究方向。

3)目前已经存在许多适用于不同类型图像的局部图像描述符,但是还没有任何一种局部图像描述符能够适用于所有类型的图像,他们对图像数据具有依赖性。因此,对各种优秀的局部图像描述符进行深入分析,取长补短,进行适当的结合,发展适应多种类型图像的稳健局部描述符将是今后局部图像描述符的一个研究工作。解决途径主要有3种:

(1)在像素层局部特征中进行融合。因为不同的像素层局部特征对局部纹理会有不同的统计模式,有些是互补的。对其进行融合会提高局部描述符的独特性和对不同纹理区域的适应性。

(2)在空间编码层进行融合。局部区域中不同的局部空间会表现出不同的外观统计模式,所以在

空间编码层对信息进行融合将是一个有效的途径。

(3)在决策层进行融合。局部图像描述符决策层融合的一个解决途径为:对多个互补的局部图像描述符直接进行级联,然后利用监督式机器学习方法进行优化和维数约简,去除冗余信息。

4)单纯的颜色数据具有很大的歧义性,Kinect的问世和低廉的售价让同时获取深度和彩色图像变得很容易。颜色(灰度)信息和3D几何空间具有内在的互补性,而这两种信息的处理方式安全不同,所以给联合描述带来了一定的困难。目前,局部RGB-D描述符的研究刚刚起步,对颜色信息和3D几何的联合描述还停留在较低层次,对局部区域的颜色信息和空间信息进行高效描述还需要进行大量的研究。一个可行的解决方法为:在处理3D点云时加入颜色信息,实现对这两种信息的深度融合。

另外,RGB-D局部空时描述符也是基于RGB-D传感器的行为识别领域的一个研究重点。RGB-D局部空时描述符包含时间、3D几何空间以及颜色3种不同类型的信息,融合这些信息并对其进行高效描述是研究的重点。因为其包含的信息太多,所以RGB-D局部空时描述符的研究难度也较大。基于RGB-D传感器的行为识别是目前的研究热点,采用局部描述符是目前行为识别主流的解决方法,所以RGB-D局部空时描述符是有效解决该问题的关键技术。

5)目前的局部图像描述符对强纹理目标效果较好,但在弱纹理目标上的局部区域描述是一个难点,加强对局部区域描述的数学统计理论研究,有效解决该问题将成为未来局部图像描述符的一个研究方向。相对于强纹理,弱纹理的像素层局部特征辨别性很弱,所以可以考虑引入区域形状,多区域节点连接关系等元素进行描述。

参考文献 (References)

- [1] Furukawa Y, Ponce J. Accurate, dense, and robust multiview stereopsis[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2010, 32(8):1362-1376.
- [2] Agarwal S, Snavely N, Simon I, et al. Building rome on a day [C]//Proceedings of 12th International Conference on Computer Vision. Kyoto: IEEE, 2009: 368-381.
- [3] Zhang G F, Jia J Y, Wong T T, et al. Consistent depth maps recovery from a video sequence[J]. IEEE Trans. on Pattern Anal-

- ysis and Machine Intelligence, 2009, 31(6):974-988.
- [4] Nister D, Stewenius H. Scalable recognition with a vocabulary tree[C]//Proceedings of Conference on Computer Vision and Pattern Recognition. New York, USA; IEEE, 2006: 2161-2168.
 - [5] Wang J, Yang J, Yu K, et al. Locality constrained linear coding for image classification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA; IEEE, 2010: 3360-3367.
 - [6] Zhang J G, Marszalek M, Lazebnik S, et al. Local features and kernels for classification of texture and object categories: a comprehensive study [J]. International Journal of Computer Vision, 2007, 73(2):213-238.
 - [7] Lazebnik S, Schmid C, Ponce J. A sparse texture representation using local affine regions [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2005, 27(8):1265-1278.
 - [8] Lowe D G. Distinctive image features from scale-invariant keypoints [J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
 - [9] Jegou H, Douze M, Schmid C. Improving bag-of-features for large scale image search [J]. International Journal of Computer Vision, 2010, 87(3):316-336.
 - [10] Jegou H, Douze M, Schmid C. Product quantization for nearest neighbor search [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2011, 33(1):117-128.
 - [11] Jegou H, Perronnin F, Douze M, et al. Aggregating local image descriptors into compact codes [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2012, 34(9):1704-1716.
 - [12] Se S, Lowe D G, Little G J. Vision-based global localization and mapping for mobile robots [J]. IEEE Transactions on Robotics, 2005, 21(3):364-375.
 - [13] Angeli A, Filliat D, Doncieux S. A fast and incremental method for loop-closure detection using bags of visual words [J]. IEEE Transactions on Robotics, 2008, 24(5):1027-1037.
 - [14] Cummins M, Newman P. Appearance-only SLAM at large scale with FAB-MAP 2.0 [J]. International Journal of Robotics Research, 2011, 30(9):1100-1123.
 - [15] Brown M, Lowe D G. Automatic panoramic image stitching using invariant features [J]. International Journal of Computer Vision, 2007, 74(1):59-73.
 - [16] Mikolajczyk K, Schmid C. A performance evaluation of local descriptors [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2005, 27(10):1615-1630.
 - [17] Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2002, 24(4):509-522.
 - [18] Schaffalitzky F, Zisserman A. Multi-View matching for unordered image sets[C]//Proceedings of the 7th European Conference on Computer Vision. Berlin Heidelberg: Springer, 2002:414-431.
 - [19] Freeman W, Adelson E. The Design and use of steerable filters [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1991, 13(9): 891-906.
 - [20] Lazebnik S, Schmid C, Ponce J. Sparse texture representation using affine-invariant neighborhoods[C]//Proceedings of Conference on Computer Vision and Pattern Recognition. Wisconsin, USA; IEEE, 2003:319-324.
 - [21] Ke Y, Sukthankar R. PCA-SIFT: a more distinctive representation for local image descriptors[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA; IEEE, 2004:511-517.
 - [22] Ce L, Yuen J, Torralba A. SIFT Flow: dense correspondence across scenes and its applications [J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2011, 33(5):978-994.
 - [23] Bay H, Ess A, Tuytelaars T, et al. SURF: speeded up robust features [J]. Computer Vision and Image Understanding, 2008, 110(3):346-359.
 - [24] Tola E, Lepetit V, Fua P. DAISY: an efficient dense descriptor applied to wide-baseline stereo [J]. IEEE Transactions on Pattern Analysis and Machine, 2010, 32(5):815-830.
 - [25] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Diego, USA; IEEE, 2005: 886-893.
 - [26] Takacs G, Chandrasekhar V, Tsai S, et al. Unified real-time tracking and recognition with rotation invariant fast features [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA; IEEE, 2010: 934-941.
 - [27] Ambai M, Yoshida Y. CARD: compact and real-time descriptors [C]//Proceedings of IEEE International Conference on Computer Vision. Barcelona, Spain; IEEE, 2011:97-104.
 - [28] Chandrasekhar V, Takacs G, Chen D. Compressed histogram of gradients: a low-bitrate descriptor [J]. International Journal of Computer Vision, 2012, 96(3): 384-399.
 - [29] Boix X, Gygli M, Roig G, et al. Sparse quantization for patch description [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA; IEEE, 2013: 2842-2849.
 - [30] Huang C R, Chen C R, Chung P C. Contrast context histogram: an efficient discriminating local descriptor for object recognition and image matching [J]. Pattern Recognition, 2008, 41(10): 3071-3077.
 - [31] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns [J]. IEEE Transactions on Pattern Analysis and Machine, 2002, 24(7):971-987.
 - [32] Ahonen T, Hadid A, Pietikainen M. Face description with local binary patterns: Application to face recognition [J]. IEEE

- Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(12):2037-2041.
- [33] Heikkilä M, Pietikainen M. A texture-based method for modeling the background and detecting moving objects [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(4):657-662.
- [34] Zhao G, Pietikainen M. Dynamic texture recognition using local binary patterns with an application to facial expressions [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6):915-928.
- [35] Wu J X, Rehg J M. CENTRIST: a visual descriptor for scene categorization [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(8):1489-1501.
- [36] Guo Z, Zhang L, Zhang D. A completed modeling of local binary pattern operator for texture classification [J]. IEEE Trans. on Image Process, 2010, 19(6):1657-1663.
- [37] Tan X, Triggs B. Enhanced local texture feature sets for face recognition under difficult lighting conditions [J]. IEEE Trans. on Image Process, 2010, 19(6):1635-1650.
- [38] Zhang B C, Gao Y S, Zhao S Q, et al. Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor [J]. IEEE Trans. on Image Process, 2010, 19(2):533-544.
- [39] Ramirez R A, Rojas C J, Chae O. Local directional number pattern for face analysis; face and expression recognition [J]. IEEE Trans. on Image Processing, 2012, 21(5):1740-1752.
- [40] Liao S, Law M W K, Chung A C S. Dominant local binary patterns for texture classification [J]. IEEE Trans. on Image Process, 2009, 18(5):1107-1118.
- [41] Zhao Y, Huang D S, Jia W. Completed local binary count for rotation invariant texture classification [J]. IEEE Trans. on Image Process, 2012, 21(10):4492-4497.
- [42] Murala S, Maheshwari R P, Balasubramanian R. Local tetra patterns: a new feature descriptor for content-based image retrieval [J]. IEEE Trans. on Image Process, 2012, 21(5):2874-2886.
- [43] Maani R, Kalra S, Yang Y H. Rotation invariant local frequency descriptors for texture classification [J]. IEEE Trans. on Image Process, 2013, 22(6):2409-2419.
- [44] Zhao G Y, Ahonen T, Matas J, et al. Rotation-invariant image and video description with local binary pattern features [J]. IEEE Trans. on Image Process, 2012, 21(4):1465-1477.
- [45] Ojansivu V, Heikkilä J. Blur insensitive texture classification using local phase quantization [C]//Proceedings of the International Conference on Image and Signal Processing. Berlin Heidelberg: Springer, 2008:236-243.
- [46] Ahonen T, Rahtu E, Ojansivu V, et al. Recognition of blurred faces using local phase quantization [C]//Proceedings of 19th International Conference on Pattern Recognition. Tampa, USA: Institute of Electrical and Electronics Engineers Inc, 2008:1-4.
- [47] Chan C H, Tahir M A, Kittler J, et al. Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(5):1164-1177.
- [48] Heikkilä M, Pietikainen M, Schmid C. Description of interest regions with local binary patterns [J]. Pattern Recognition, 2009, 42(3):425-436.
- [49] Gupta R, Patil H, Mittal A. Robust order-based methods for feature description [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA: IEEE, 2010:334-341.
- [50] Zeng H, Mu Z C, Wang X Q. A robust method for local image feature region description [J]. Acta Automatica Sinica, 2011, 37(6):658-664. [曾慧, 穆志纯, 王秀青. 一种鲁棒的图像局部特征区域的描述方法 [J]. 自动化学报, 2011, 37(6):658-664]
- [51] Chen J, Shan S G, He C, et al. WLD: a robust local image descriptor [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9):1705-1720.
- [52] Gong D Y, Li S T, Xiang Y. Face recognition using weber local descriptors [C]//Proceedings of First Asian Conference on Pattern Recognition. Beijing, China: IEEE, 2011:589-592.
- [53] Hussain M, Muhammad G, Bebis G. Face recognition using multiscale and spatially enhanced weber law descriptor [C]//Proceedings of the 8th International Conference on Signal Image Technology and Internet Based Systems. Naples, Italy: IEEE, 2012:85-89.
- [54] Muhammad G, Hussain M, Alenezy F, et al. Race recognition from face images using Weber local descriptor [C]//Proceedings of the 19th International Conference on Systems, Signals and Image Processing. Vienna, Austria: IEEE, 2012:421-424.
- [55] Ullah I, Hussain M, Muhammad G. Gender recognition from face images with local WLD descriptor [C]//Proceedings of the 19th International Conference on Systems, Signals and Image Processing. Vienna, Austria: IEEE, 2012:417-420.
- [56] Gupta R, Mittal A. SMD: a locally stable monotonic change invariant feature descriptor [C]//Proceedings of the 10th European Conf. Computer Vision. Berlin Heidelberg: Springer, 2008:265-277.
- [57] Gupta R, Mittal A. Illumination and affine-invariant point matching using an ordinal approach [C]//Proceedings of IEEE 11th International Conference on Computer Vision. Rio de Janeiro, Brazil: IEEE, 2007:1-8.
- [58] Tang F, Lim S H, Change N L, et al. A novel feature descriptor invariant to complex brightness changes [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009:2631-2638.
- [59] Fan B, Wu F C, Hu Z Y. Rotationally invariant descriptors using

- intensity order pooling [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(10): 2031-2045.
- [60] Wang Z H, Fan B, Wu F C. Local intensity order pattern for feature description [C]//Proceedings of IEEE International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011: 603-610.
- [61] Kim B, Yoo H, Sohn K. Exact order based feature descriptor for illumination robust image matching [J]. Pattern Recognition, 2013, 46(12): 3268-3278.
- [62] Calonder M, Lepetit V, Strecha C, et al. Brief: binary robust independent elementary features [C]//Proceedings of the 11th European conference on Computer vision. Berlin Heidelberg: Springer, 2010: 778-792.
- [63] Calonder M, Lepetit V, Ozuysal M, et al. BRIEF: computing a local binary descriptor very fast [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(7): 1281-1298.
- [64] Galvezlopez D, Tardos J D. Bags of binary words for fast place recognition in image sequences [J]. IEEE Transactions on Robotics, 2012, 28(5): 1188-1197.
- [65] Rublee E, Rabaud V, Konolige K, et al. ORB: an efficient alternative to SIFT or SURF [C]//Proceedings of International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011: 2564-2571.
- [66] Leutenegger S, Chli M, Siegwart R. BRISK: binary robust invariant scalable keypoints [C]//Proceedings of the IEEE International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011: 2548-2555.
- [67] Heinly J, Dunn E, Frahm J M. Comparative evaluation of binary features [C]//Proceedings of the 12th European Conference on Computer Vision. Berlin Heidelberg: Springer, 2012: 7-13.
- [68] Alahi A, Ortiz R, Vandergheynst P. FREAK: fast retina keypoint [C]//Proceedings of the IEEE International Conference on Computer Vision. Providence, USA: IEEE, 2012: 510-517.
- [69] Ziegler A, Christiansen E, Kriegman D, et al. Locally uniform comparison image descriptor [C]//Proceedings of Neural Information Processing Systems. Nevada, USA: Neural Information Processing System Foundation, 2012: 1-9.
- [70] Bo L, Ren X, Fox D. Kernel descriptors for visual recognition [C]//Proceedings of Neural Information Processing Systems. Vancouver, Canada: Curran Associates Inc, 2010: 244-252.
- [71] Wang P, Wang J D, Zeng G, et al. Supervised kernel descriptor for visual recognition [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA: IEEE, 2013: 2858-2865.
- [72] Ren X, Bo L, Fox D. Rgb-(d) scene labeling: Features and algorithms [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Providence, USA: IEEE, 2012: 2759-2766.
- [73] Natarajan P, Wu S, Vitaladevuni S N P, et al. Multi-channel shape-flow kernel descriptors for robust video event detection and retrieval [C]//Proceedings of the 12th European Conference on Computer Vision. Berlin Heidelberg: Springer, 2012: 301-314.
- [74] Tuzel O, Porikli F, Meer P. Region covariance: a fast descriptor for detection and classification [C]//Proceedings of 9th European Conf. Computer Vision. Berlin Heidelberg: Springer, 2006: 589-600.
- [75] Tuzel O, Porikli F, Meer P. Pedestrian detection via classification on riemannian manifolds [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(10): 1713-1727.
- [76] Pang Y W, Yuan Y, Li X L. Gabor-based region covariance matrices for face recognition [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2008, 18(7): 989-993.
- [77] Paisitkriangkrai S, Shen C H, Zhang J. Fast pedestrian detection using a cascade of boosted covariance features [J]. IEEE Trans. on Circuits and Systems for Video Technology, 2008, 18(8): 1140-1151.
- [78] Heikkila J. Pattern matching with affine moment descriptors [J]. Pattern Recognition, 2004, 37(9): 1825-1834.
- [79] Zhang D, Lu G. Review of shape representation and description techniques [J]. Pattern Recognition, 2004, 37(1): 1-19.
- [80] Chen Z, Sun S K A. Zernike Moment Phase-Based Descriptor for Local Image Representation and Matching [J]. IEEE Trans. on Image Process, 2010, 19(1): 205-219.
- [81] Khotanad A, Hong Y H. Invariant image recognition by Zernike moments [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1990, 12(5): 489-497.
- [82] Winder S, Hua G, Brown M. Picking the best DAISY [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009: 178-185.
- [83] Brown M, Hua G, Winder S. Discriminative learning of local image descriptors [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(1): 43-57.
- [84] Simonyan K, Vedaldi A, Zisserman A. Learning local feature descriptors using convex optimisation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(8): 1573-1585.
- [85] Trzcinski T, Christoudias M, Lepetit V, et al. Learning image descriptors with the boosting-trick [C]//Proceedings of Neural Information Processing Systems. Nevada, USA: Neural Information Processing System Foundation, 2012: 278-286.
- [86] Mikolajczyk K, Matas J. Improving descriptors for fast tree matching by optimal linear projection [C]//Proceedings of IEEE International Conference on Computer Vision. Rio de Janeiro, Brazil: IEEE, 2007: 1-8.
- [87] Hua G, Brown M, Winder S. Discriminant embedding for local image descriptors [C]//Proceedings of IEEE International Conference on Computer Vision. Rio de Janeiro, Brazil: IEEE,

- 2007;1-8.
- [88] Cai H P, Mikolajczyk K, Matas J. Learning linear discriminant projections for dimensionality reduction of image descriptors [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 33(2): 338-352.
- [89] Xu X M, Yang D, Zhang X H, et al. The algorithm of descriptor based on locality preserving projections [J]. Acta Automatica Sinica, 2008, 34(9): 1174-1177. [徐小明, 杨丹, 张小洪, 等. 基于局部不变映射的特征描述器算法 [J]. 自动化学报, 2008, 34(9): 1174-1177]
- [90] Strecha C, Bronstein A M, Bronstein M M, et al. LDAHash: improved matching with smaller descriptors [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(1): 66-78.
- [91] Trzcinski T, Lepetit V. Efficient discriminative projections for compact binary descriptors [C]//Proceedings of the 12th European Conference on Computer Vision. Florence, Italy: Springer Berlin Heidelberg, 2012;7-13.
- [92] Trzcinski T, Christoudias M, Fua P, et al. Boosting binary key-point descriptors [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA: IEEE, 2013: 2874-2881.
- [93] Bosch A, Zisserman A, Munoz, X. Scene classification using a hybrid generative/discriminative approach [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(4): 712-727.
- [94] Sande K E A, Gevers T, Snoek C G M. Evaluating color descriptors for object and scene recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(9): 1582-1596.
- [95] Burghouts G J, Geusebroek J M. Performance evaluation of local colour invariants [J]. Computer Vision and Image Understanding, 2009, 113(1): 48-62.
- [96] Weijer J, Gevers T, Smeulders A W. Robust photometric invariant features from the color tensor [J]. IEEE Trans. on Image Processing, 2006, 15(1): 118-127.
- [97] Brown M, Susstrunk S. Multi-spectral SIFT for scene category recognition [C]//Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition. Providence, RI: IEEE, 2011: 177-184.
- [98] Weijer J, Gevers T, Bagdanov A D. Boosting color saliency in image feature detection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(1): 150-156.
- [99] Zhang J, Barhomi Y, Thomas S. A new biologically inspired color image descriptor [C]//Proceedings of the 12th European Conference on Computer Vision. Florence, Italy: Springer Berlin Heidelberg, 2012;7-13.
- [100] Blum M, Springenberg J T, Wulff J, et al. A learned feature descriptor for object recognition in RGB-D data [C]//Proceedings of IEEE International Conference on Robotics and Automation. Saint Paul, USA: IEEE, 2012: 1298-1303.
- [101] Bo L F, Lai K V, Ren X F, et al. Object Recognition with Hierarchical Kernel Descriptors [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Providence, RI: IEEE, 2011;1729-1736.
- [102] Bo L F, Ren X F, Fox D. Depth kernel descriptors for object recognition [C]//Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. San Francisco, USA: IEEE, 2011;821-826.
- [103] Tombari F, Salti S, Stefano L D. A combined texture-shape descriptor for enhanced 3D feature matching [C]//Proceedings of the 18th IEEE International Conference on Image Processing. Brussels: IEEE, 2011;809-812.
- [104] Tombari F, Salti S D, Stefano L. Unique signatures of histograms for local surface description [C]//Proceedings of European Conference on Computer Vision. Berlin Heidelberg: Springer, 2010;356-369.
- [105] Nascimento E R, Oliveira G L, Campos M F M, et al. BRAND: a robust appearance and depth descriptor for RGB-D images [C]//Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. Vilamoura, Portugal: IEEE, 2012;1720-1726.
- [106] Laptev I. On space-time interest points [J]. International Journal of Computer Vision, 2005, 64(2/3): 107-123.
- [107] Natarajan P, Wu S, Vitaladevuni S N P, et al. Multi-channel shape-flow kernel descriptors for robust video event detection and retrieval [C]//Proceedings of the 12th European Conference on Computer Vision. Berlin Heidelberg: Springer, 2012: 301-314.
- [108] Trulls E, Sanfeliu A, Moreno-Noguer F. Spatiotemporal descriptor for wide-baseline stereo reconstruction of non-rigid and ambiguous scenes [C]//Proceedings of the 12th European Conference on Computer Vision. Berlin Heidelberg: Springer, 2012;441-454.
- [109] Laptev I, Marszałek M, Schmid C, et al. Learning realistic human actions from movies [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA: IEEE, 2008;1-8.
- [110] Scovanner P, Ali S, Shah M. A 3-dimensional sift descriptor and its application to action recognition [C]//Proceedings of International conference on Multimedia. Augsburg, Germany: Association for Computing Machinery, 2007: 24-29.
- [111] Klaser A, Marszałek M, Schmid C. A spatio-temporal descriptor based on 3D-gradients [C]//Proceedings of the British Machine Vision Conference. Leeds, UK: British Machine Vision Association, 2008;275-284.
- [112] Zhao G, Pietikainen M. Dynamic texture recognition using volume local binary patterns [C]//Proceedings of Workshop on Dy-

- namical Vision. Berlin: Springer, 2006:165-177.
- [113] Willems G, Tuytelaars T, Gool L V. An efficient dense and scale-invariant spatio-temporal interest point detector[C]//Proceedings of the 10th European Conference on Computer Vision. Berlin Heidelberg: Springer, 2008: 650-663.
- [114] Zhao G, Pietikainen M. Dynamic texture recognition using local binary patterns with an application to facial expressions [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(6):915-928.
- [115] Chen J, Zhao G, Pietikainen M. An improved local descriptor and threshold learning for unsupervised dynamic texture segmentation [C]//Proceedings of IEEE International Conference on Computer Vision Workshops. Kyoto, Japan: IEEE, 2009:460-467.
- [116] Paivarinta J, Rahtu E, Heikkila J. Volume local phase quantization for blur-insensitive dynamic texture classification [C]//Proceedings of Scandinavian conference on Image analysis. Berlin: Springer Verlag, 2011: 360-369.
- [117] Norouznezhad E, Harandi M T, Bigdeli A, et al. Directional space-time oriented gradients for 3D visual pattern analysis [C]//Proceedings of the 12th European Conference on Computer Vision. Berlin Heidelberg: Springer, 2012: 736-749.
- [118] Xia L, Aggarwal J K. Spatio-temporal depth cuboid similarity feature for activity recognition using depth camera [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Portland, USA: IEEE, 2009: 2834-2841.
- [119] Zhang H, Parker L E. 4-Dimensional local spatio-temporal features for human activity recognition [C]//Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. San Francisco, USA: IEEE, 2011: 2044-2049.
- [120] Wu J X, Rehg J M. Beyond the euclidean distance: creating effective visual codebooks using the histogram intersection kernel [C]//Proceedings of IEEE International Conference on Computer Vision. Kyoto, Japan: IEEE, 2009: 630-637.
- [121] Pele O, Werman M. Fast and robust earth mover's distances [C]//Proceedings of IEEE International Conference on Computer Vision. Kyoto, Japan: IEEE, 2009:460-467.
- [122] Jegou H, Douze M, Schmid C. Improving Bag-of-Features for large scale image search [J]. International Journal of Computer Vision, 2010, 87(3): 316-336.
- [123] Jegou R, Douze M, Schmid C. Product quantization for nearest neighbor search [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(1): 117-128.
- [124] Sivic J, Zisserman A. Video google: a text retrieval approach to object matching in videos [C]//Proceedings of the 9th IEEE International Conference on Computer Vision. Nice, France: IEEE, 2003: 1470-1477.
- [125] Gemert J C, Geusebroek J M, Veenman C J, et al. Kernel codebooks for scene categorization [C]//Proceedings of the European Conference on Computer Vision. Berlin Heidelberg: Springer, 2008: 696-709.
- [126] Perronnin F S, Nchez J, Mensink T. Improving the fisher kernel for large-scale image classification [C]//Proceedings of the European Conference on Computer Vision. Berlin Heidelberg: Springer, 2010: 143-156.
- [127] Jegou H, Perronnin F, Douze M, et al. Aggregating local image descriptors into compact codes [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(9): 1704-1716.
- [128] Zhou X, Yu K, Zhang T, et al. Image classification using super-vector coding of local image descriptors [C]//Proceedings of the European Conference on Computer Vision. Berlin Heidelberg: Springer, 2010: 141-154.
- [129] Moreels P, Perona P. Evaluation of features detectors and descriptors based on 3D objects [J]. International Journal of Computer Vision, 2007, 73(3):263-284.
- [130] Aanæs H, Dahl A, Pedersen K S. Interesting interest points: a comparative study of interest point performance on a unique data set [J]. International Journal of Computer Vision, 2012, 97(1):18-35.
- [131] Kaneva B, Torralba A, Freeman W T. Evaluation of image features using a photorealistic virtual world [C]//Proceedings of IEEE International Conference on Computer Vision. Barcelona, Spain: IEEE, 2011:2282-2289.
- [132] Gauglitz S, Hollerer T, Turk M. Evaluation of interest point detectors and feature descriptors for visual tracking [J]. International Journal of Computer Vision, 2011, 94(3) 335-360.
- [133] Gil A, Mozos O M, Ballesta M, et al. A comparative evaluation of interest point detectors and local descriptors for visual SLAM [J]. Machine Vision and Applications, 2010, 21(6):905-920.