软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2016,27(10):2462−2472 [doi: 10.13328/j.cnki.jos.005090]
©中国科学院软件研究所版权所有.
E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

层次结构 K-d 树的立体图像快速匹配方法*

张贵安, 袁志勇, 童倩倩, 廖祥云

(武汉大学 计算机学院,湖北 武汉 430072)

通讯作者: 袁志勇, E-mail: zhiyongyuan@whu.edu.cn

摘 要: 特征匹配是计算机视觉和图形图像处理领域中很多研究方向的基础,也是当前的研究热点.SIFT(scale-invariant feature transformation)特征因其具有尺度、旋转不变性,对一定范围的仿射及视角变换具有鲁棒性等优点,自 Lowe 提出后,10 多年来一直受到众多研究人员的关注.匹配的快速性和准确性是很多应用对特征匹配的要求,如三维重建中立体图像对(stereo pairwise image,简称 SPI)的匹配.针对这一问题,以 SIFT 特征为基础,提出用于 SPI 匹配的方向大约一致(approximately consistent in orientation,简称 ACIO)约束关系,其描述了 SPI 的匹配特征向量间的空间位置关系,有效地避免了误匹配的发生,提高了匹配的精度;通过对标准 K-d 树(standard K-d tree,简称 SKD-tree)结构的分析,提出了层次结构 K-d 树(hierarchical K-d tree,简称 HKD-tree),将 SPI 特征集根据 ACIO 约束关系划分成层次结构并建立映射,该方法缩小了搜索空间,从而达到加速匹配的目的.在 ACIO 和 HKD-tree 的基础上,提出了高效、快速的匹配算法.实验结果表明,所提方法比 SKD-tree 方法和最新的级联哈希方法(cascade hash,简称 CasHash)在精度上略占优势,但在匹配速度上比 SKD-tree 快一个数量级以上,同时也数倍于 CasHash.

关键词: 尺度不变特征变换;方向大约一致;层次结构 K-d 树;立体图像对
中图法分类号: TP391

中文引用格式: 张贵安,袁志勇,童倩倩,廖祥云.层次结构 K-d 树的立体图像快速匹配方法.软件学报,2016,27(10):2462−2472. http://www.jos.org.cn/1000-9825/5090.htm

英文引用格式: Zhang GA, Yuan ZY, Tong QQ, Liao XY. Fast and hierarchical K-d tree based stereo image matching method. Ruan Jian Xue Bao/Journal of Software, 2016,27(10):2462−2472 (in Chinese). http://www.jos.org.cn/1000-9825/5090.htm

Fast and Hierarchical K-d Tree Based Stereo Image Matching Method

ZHANG Gui-An, YUAN Zhi-Yong, TONG Qian-Qian, LIAO Xiang-Yun

(School of Computer, Wuhan University, Wuhan 430072, China)

Abstract: Feature Matching has long been the basis and a central topic in the field of computer vision and image processing. SIFT (scale-invariant feature transformation, by David G. Lowe), due to its advantages of invariance to image scale and rotation, and robustness to a substantial range of affine distortion and change in viewpoint, has been attracting the attention of many domestic and foreign researchers over a decade. Rapidity and accuracy are very crucial for stereo pairwise image matching in applications such as 3D reconstruction. First, in order to accelerate the speed and promote the accuracy of matching, this paper proposes a novel method based on SIFT called approximately consistent in orientation (ACIO), which depicts the spatial location relationship of two matched vectors between stereo pairwise images (SPI), and therefore improves the accuracy of matching efficiently by avoiding the wrong correspondences. Secondly, this paper analyzes the structure of standard K-d tree (SKD-tree) and proposes a new one with hierarchical structure, named HKD-tree, which partitions the feature sets of SPI into stripes in terms of ACIO constraint and builds map between them. By reducing the search space, the matching speed increases greatly. Thirdly this work presents an efficient and fast matching algorithm based on ACIO and HKD-tree. Extensive trials based on a benchmark data set show that the proposed approach outperforms the

* 基金项目: 国家自然科学基金(61373107)
Foundation item: National Natural Science Foundation of China (61373107)
收稿时间: 2016-02-02; 修改时间: 2016-03-25; 采用时间: 2016-05-09; jos 在线出版时间: 2016-08-08
CNKI 网络优先出版: 2016-08-09 15:37:56, http://www.cnki.net/kcms/detail/11.2560.TP.20160809.1537.008.html

© 中国科学院软件研究所 http://www.jos.org.cn

state-of-the-art methods in matching speed with slight promotion in accuracy. Particulary, it is one order of magnitude faster than SKD-tree and also several times against the recent CasHash method.

Key words: SIFT (scale-invariant feature transformation); ACIO (approximately consistent in orientation); HKD-tree; SPI (stereo pairwise image)

众所周知,视觉是灵长类动物(primate)感知周围世界的重要手段,例如人类的双目视觉为我们提供了约80%以上的外部信息,远多于其他生理器官.计算机视觉的诞生正是为了让机器拥有同样的功能,以满足人类的生活生产需要,如视频异常监控、机器导航、生产线缺陷检测、医疗定位、三维重建等应用都需要视觉辅助.在这些应用实现中,最重要且花费时间最多的环节就是图像匹配,探索高精度、高效率的匹配方法是当前计算机视觉领域研究的热点和难点,本文针对这一目标展开了研究.

通常对同一对象或场景进行拍摄时,会存在视角、尺度、旋转等的不同,这给图像匹配带来了很大的挑战;而在一些使用立体图像对(stereo pairwise image,简称 SPI)的应用中,会有尺度变化和连续小角度的视角变化,没有平面内的旋转变化(rotation-in-plain).如图 1 所示,左右图像对虽然存在着尺度上的较大差异,但是通过观察发现,如果在左图像上的两个特征点的位置关系是左右的,那么它们的匹配点在右图像上的位置也是左右关系,即空间位置没有发生本质的变化;当将匹配的特征点对用直线连接起来时,我们发现这些匹配线基本上都指向同一个方向.基于这一发现,我们提出了基于空间位置关系的方向大约一致(approximately consistent in orientation,简称 ACIO)约束关系,该约束可以很好地避免特征点的误匹配,同时也激发了我们对标准 K-d 树(SKD-tree)改进的灵感.SKD-Tree 是众多树结构中匹配效率最高的一种结构,时间复杂度为 $O(\log(N))$ (N 为特征点个数),但是随着特征点个数的急剧增多,这种优势将不复存在.本文在 SKD-tree 的基础上提出了层次结构 K-d 树(HKD-tree),不仅保持了 SKD-tree 的优点,而且在整体的匹配速率上得到了很大的提升.随后,在 ACIO 和 HKD-tree 的基础上,提出了快速、精确的匹配算法.

Fig.1 Spatial position relationship of features and orientation of matched feature vectors

图 1 特征点的空间位置关系及匹配特征向量的方向性

为了验证本文所提方法的有效性,对基于 SKD-tree 的方法(也称为标准方法)和最新方法在标准数据库上做了对比实验,实验结果表明,本文方法不仅在精度上比其他方法更精确,在匹配速度上比标准方法快一个数量级以上,同时比最新方法快数倍.匹配速度的大幅度提升,将对诸如快速三维重建等应用产生很大的促进作用.

综上,本文的主要贡献有:

(1) 提出基于空间位置关系的方向大约一致(ACIO)约束关系,并给出具体的定义;

© 中国科学院软件研究所 http://www.jos.org.cn

(2) 提出对 SKD-tree 改进的层次结构 K-d 树,即 HKD-tree;

(3) 在 ACIO 约束下,提出用于 HKD-tree 的匹配算法;

(4) 通过大量的实验验证了本文方法的有效性,同时分析了其可扩展性,为进一步的研究作准备.

本文第 1 节是相关工作,主要介绍特征匹配方法的发展和应用.第 2 节对本文所提方法作详细的介绍,包括约束关系的定义、结构的描述和算法的设计.第 3 节通过对比实验来验证和分析本文方法的性能.第 4 节总结分析了本文方法的局限性和未来的研究重点和方向.

1 相关工作

在计算机视觉和图形图像处理领域中,图像匹配是一个基础且至关重要的环节,通过提取图像中的特征并计算各个特征间的相似度来判断是否匹配,从而达到如识别、分类等目的,因此特征类型的选择非常关键.图像特征从尺度上分为全局特征和局部特征,全局特征包括颜色、纹理、形状等,局部特征包括点、线、区域等,由于局部特征具有较强的稳定性,不易受外界干扰,尤其是点特征,对光照和仿射变化具有一定的鲁棒性[1],得到了研究人员的广泛关注,并提出了很多关于局部特征描述的方法.方向梯度直方图(histogram of oriented gradient,简称 HOG)特征描述子[2],通过计算局部区域的梯度方向并统计,形成直方图,用此来表示一个特征.但是,由于梯度的性质,导致其对噪声非常敏感;Lowe 在 1999 年[3]初次提出了尺度不变特征变换(scale invariant features transform,简称 SIFT),并在 2004 年[4]对此作了更为详细的阐述和应用的介绍,SIFT 特征对图像尺度、旋转、一定角度范围的仿射和视角等变化具有不变性,同时对于一些噪声和光照变化也有一定的鲁棒性,在图像检索[5,6]、物体检测[7]和识别[8,9]、场景分类[10,11]等方面得到了成功的应用.

由于 SIFT 在诸多应用中的出色表现,研究人员对此展开了大量的研究,总体上可以分为 3 类.

(1) 与 SIFT 功能类似(SIFT-like)的特征描述符:SURF(speeded up robust features)是由 Bay 等人[12]提出,在兴趣点探测阶段采用了不同于 SIFT 在 DoG 空间检测极值点的方式,而是利用积分图(integral image)在盒式卷积上计算的快速性,通过近似计算海森矩阵(Hessian matrix)的行列式并判断其是否为极大值来选择兴趣点.描述符计算的过程和 SIFT 类似,但是用 x 和 y 方向的哈尔小波响应(Haar wavelet response)代替了梯度,因为这可以继续发挥积分图的计算优势;SIFT 是通过计算基于梯度的描述符之间的欧式距离来比较相似度,而 SIFT-Rank[13]是先将描述符转换成按照每个分量进行排序,然后取其序号作为新元素的顺序描述符(ordinal descriptor),最后再计算 Spearman 相关系数,将欧氏距离的平方映射到[−1,1]之间.虽然匹配的精确度有所提高,但其在描述符转换时需要 $O(N\log N)$ 的时间,还要额外的距离映射时间,相比 SIFT 而言,时间复杂度较高;Calonder 等人[14]提出的 BRIEF(binary robust independent elementary feature)是采用二进制字符串的方式构造描述符,匹配时计算汉明距离(Hamming distance)来获得相似度.由于汉明距离可以通过计算异或操作(XOR operation)来完成,所以匹配速度很快,但是 BRIEF 的主要缺陷是不具有旋转不变性.正因为如此,Rublee 等人[15]在 BRIEF 的基础上进行了改进,提出了 ORB(oriented FAST and rotated BRIEF)方法,克服了 FAST 的无方向性和 BRIEF 的无旋转不变性,取得了不错的实验结果.LI 等人[16]也针对 BRIEF 的缺点进行了改进,提出具有旋转不变性的 RIBRIEF(rotaion Invariant BRIEF)描述符,可是该方法和前面两种基于 BRIEF 的方法都需要对特征进行学习.

(2) 对 SIFT 的优化:如对度量方法的优化,Pele 和 Werman[17]采用了有别于传统欧式距离的陆地移动距离 EMD(earth mover’s distance).类似地,还有 diffusion distance[18]和 $\mathrm{EMD}_{\mathrm{MOD}}$[19].对描述符结构的优化,主要是优化描述符的长度问题(标准的长度是 128 位),Yan 和 Sukthankar[20]提出的 PCA-SIFT 是将主成分分析技术运用到描述符维度的优化上,在以特征点为中心的 41×41 的图像块上计算 2×39×39=3042 维的向量,再运用 PCA 技术将维度降到 36 维.虽然在匹配速度上有所提升,但是 Mikolajczyk 和 Schmid[21]已经证明其在独特性(distinctiveness)上不如 SIFT,同时还会降低特征点的计算效率.Gilinsky 和 Zelnik-Manor[22]提出了一种压缩的描述符的表示——SIFTpack,其考虑到邻近的两个兴趣点的描述符有可能存在重叠,这样的话,重叠部分就被重复存储,对这个问题的研究不仅可以降低内存消耗,同时还会带来匹配性能上的提升.最近的优化方法是由 Cheng

© 中国科学院软件研究所 http://www.jos.org.cn

等人[1]提出的基于哈希的方法,称为级联哈希(cascade Hashing),它首先用一个短编码的哈希查找来进行一遍粗略搜索,为参考图像 I 建立一个查找表,表中有多个桶(bucket),I 中的查询特征点 p 在目标图像 J 中的匹配点都会落入同一个桶中.经过粗略的搜索后,在哈希查找表上通过计算每个候选对象的汉明距离来进行精细搜索.最后在经过汉明距离排序的候选中,选择前 k 个点,再通过欧式距离获得最近邻和次近邻用于距离比值,获得匹配点.通过 Cheng 等人的实验,该方法优于 SKD-tree 方法,且比同类的哈希方法 LDAHash[23]有更高的匹配效率.这也是本文将要对比的方法之一,具体见实验部分.

(3) 对 SIFT 的扩展:Liu 等人[24]受光流的启发,提出了 SIFT Flow,在一个数据集中先用直方图相交(histogram intersection)寻找其最近邻,然后在最近邻中将查询图像(query image)通过最小化对齐能量(minimum alignment energy)来对齐于一个与之最接近的图像.这可用于在一个静态图像中预测可能的运动速度与方向,以及将运动对象传送到与之有相似场景的静态图像中,达到合成运动的目的.Spatio-temporal SIFT[25]将时空信息引入到 DoG 的计算中,并将视频帧集叠加成一个时空体[26],那么极值点的选取就在时空差分金字塔中 3 个方向(xy、xt 和 yt)的切片中进行,这样获得的特征点拥有时间属性,这有利于对视频流的一些处理,如人的肢体动作检测、视频检索等.

快速匹配算法的研究主要演化为两大类:局部的和全局的方法.局部区域的匹配算法主要是基于窗口的特征计算,如颜色和灰度等[27−31],并通过度量方法,如绝对差和(sum of absolute differences,简称 SAD)、平方差和(sum of squared differences,简称 SSD)、归一化互相关(normalized cross correlation,简称 NCC)等[32]进行比较,Banks 等人[32]通过实验发现,SAD 和 SSD 不仅原理简单且计算量小,多用于对匹配速度有较高要求的应用,而 NCC 采用了能量函数的思想,常用于尺寸较大的图像匹配,准确率较高,但其计算复杂,耗时多.为了克服光照和噪声等因素的影响,Zabih 和 Woodfill[33]提出了 CT(concensus transform)和 RT(rank transform)两种无参局部变换,其中 CT 在窗口中取某一像素为中心,并与其相邻像素灰度作比较,形成由 0 和 1 组成的位串(bit string),而 RT 取值为其中灰度小于中心像素的相邻像素个数,这两种变换相当于对图像进行了重新编码,使得相关计算与灰度值无关,从而降低了对外点(outliers)和噪声的敏感度.同时,匹配时通过计算汉明距离来衡量相似度,这可以有效地提高计算速度.全局匹配方法主要是通过优化全局能量函数($E=E_{\text{data}}+E_{\text{smooth}}$),不断地迭代平滑项以最小化能量,实现对视差的求精,目前采用的优化方法有动态规划[34,35]、图割[36,37]、置信传播[38,39].不同于从区域角度的划分,Scharstein 和 Szeliski[40]将立体匹配算法分为 4 个步骤,分别为匹配代价计算(matching cost computation)、代价聚合(cost aggregation)、视差计算(disparity computation)和精细化处理(disparity refinement).针对每一个步骤所使用的算法作了比较系统的归类,并开放了一个标准数据集[41],为后续的研究工作提供了很好的参考.

2 ACIO 和 HKD-tree 定义及算法设计

本节中,我们将对所提方法进行详细的阐述.首先是方向大约一致(ACIO)约束关系的定义,然后介绍 HKD-tree 的结构与划分方法,最后是根据前面两项内容设计快速的匹配算法.

2.1 ACIO定义

在单个物体(或场景)中,局部特征间都存在着简单的相对顺序性.在三维重建等立体视觉应用中,采集图像时需要布置多台摄像机或者通过水平移动摄像机完成多幅图像的采集,其中的两幅图像(尤其是邻近的)称为立体图像对 SPI,它们之间会有平移的位置变化以及少许的尺度缩放和视角变化.前者会使得 SPI 的重复部分减少,而后者则会导致同一幅图像中局部特征间的位置发生相对变化,但不改变顺序性.将 SPI 中的匹配点用线连接起来,我们发现这些线也存在着一定的位置关系.基于上述发现,本文提出 ACIO 约束关系,下面给出相关的定义.

- 一般地,在立体匹配中,通常将用于查询的特征,即查询特征(query feature)所在的特征点集合称为参考集(reference set,简称 RS),而被查询的特征点集合称为目标集(target set,简称 TS).因此,我们定义参考图像的特征点集合为

© 中国科学院软件研究所 http://www.jos.org.cn

$$R: R=\{r_i \mid (x_r, y_r), i\in[0, M], M\in\Re^+\},$$

而目标图像的特征点集合为

$$T: T=\{t_j \mid (x_t, y_t), j\in[0, N], N\in\Re^+\},$$

其中,M 和 N 分别是 R 和 T 特征点的数目.如果特征点 r 和 t 是匹配的,则称为匹配点对,表示为 $c(r,t)$,对应的匹配点对集合为 C;匹配点对之间的连线为匹配线 l,加上线的方向属性 Ori 后则为匹配向量,表示为

$$\boldsymbol{l}=\{\bar{c}, Ori\}, Ori\in\left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \tag{1}$$

- 定义匹配特征向量集,即由全部 P 个匹配特征向量组成的集合:

$$\boldsymbol{L}=\{\boldsymbol{l}_k \mid \langle r, t\rangle, k\in[0, P], P=\min(M, N), r\in R, t\in T\} \tag{2}$$

- 定义 ACIO 约束,如果已知两条匹配特征向量 $\boldsymbol{l}_{up}$ 和 $\boldsymbol{l}_{down}$,且它们的方向为 Ori_{up} 和 Ori_{down},那么与位于它们之间的匹配特征向量 $\boldsymbol{l}_{mid}$ 具有以下关系:

$$\boldsymbol{l}_{mid}=\alpha\boldsymbol{l}_{up}+\beta\boldsymbol{l}_{down}, \alpha, \beta\in\Re^+ \tag{3}$$

$$Ori_{mid}=\frac{Ori_{up}+Ori_{down}}{2}+\varepsilon, \varepsilon\in\left(-\frac{|Ori_{up}-Ori_{down}|}{2}, \frac{|Ori_{up}-Ori_{down}|}{2}\right) \tag{4}$$

2.2 HKD-tree设计

层次结构 K-d 树的主要思想是根据 ACIO 约束关系将 SPI 进行分层,使得相近位置的查询特征只在对应层中搜索匹配特征.具体步骤如下:

- 提取 SPI 的特征点集 R 和 T,分别取最顶端和最底端的匹配特征向量 $\boldsymbol{l}_{top}$ 和 $\boldsymbol{l}_{bottom}$(如图 2 所示,不失一般性,将尺度小的图像放在右侧),这可以应对一般情形下 SPI 可能存在的尺度、视角等不同的问题.
- 根据匹配特征向量 $\boldsymbol{l}_{top}$ 和 $\boldsymbol{l}_{bottom}$ 在特征点集 R 中的特征点 r_{top} 和 r_{bottom} 的纵坐标 $y_{r_{top}}$ 和 $y_{r_{bottom}}$,将 R 均匀地分成 H 层(无透视变换条件下),每一层 s_{rk} 称为一条条带(stripe),每一层的高度为

$$h_ref_{s_{rk}}=\frac{y_{r_{bottom}}-y_{r_{top}}}{H}(k=1, ..., H) \tag{5}$$

特征集 T 划分成同样的层数,即 H 层,并按照上述方法计算出每个条带的高度.

$$h_tar_{s_{tk}}=\frac{y_{t_{bottom}}-y_{t_{top}}}{H}(k=1, ..., H) \tag{6}$$

- 在特征提取后,先对获取的左右特征点集按照纵坐标升序排序,然后对 R 和 T 分别以各自的层高为步长,将它们划分成 H 条条带,$map_{rk\to tk}=(s_{rk}, s_{tk})(k=1, ..., H)$ 称为第 k 条 r 到 t 的条带映射.对 T 的每一条条带建立一个 K-d 树 $kdtree_k$,如图 3 所示,将其根保存在列表 $root_list$ 中,并建立 s_{tk} 和 $kdtree_k$ 的映射关系 $map_{s_{tk}\to hkdtree_k}=(s_{tk}, root_list)$.通过二级映射,建立查询特征和目标特征之间的联系

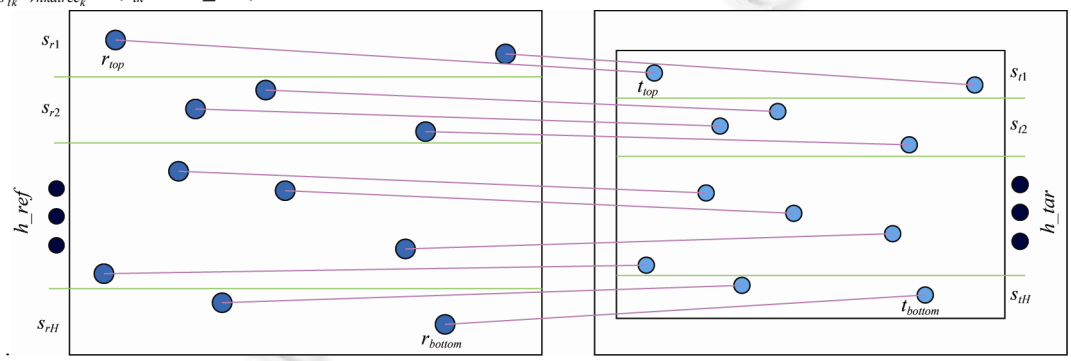

Fig.2 Hierarchical structure of partitioned SPI

图 2 SPI 的层次划分

© 中国科学院软件研究所 http://www.jos.org.cn

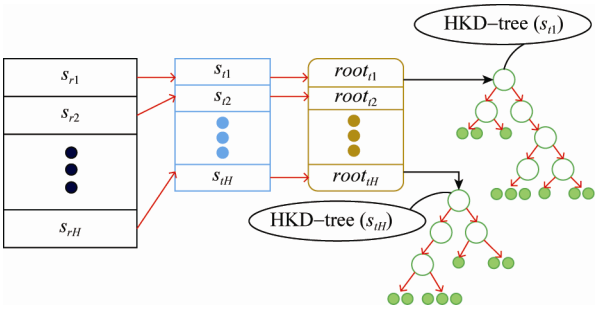

Fig.3 HKD-Tree creation and maps establishment

图 3 创建层次结构 K-d 树及建立映射关系

2.3 匹配算法设计

在 ACIO 约束关系和 HKD-tree 搜索结构的基础上,本文对基于标准 K-d 树的匹配算法进行了改进.在 HKD-tree 上,根据查询特征点(来自参考特征点集)的所属条带,由条带映射 $map_{rk\to tk}$ 找到其目标条带,即目标特征点集,然后在 *root_list* 中找到相应的树,最后在查询到的树上进行最近邻和次近邻的搜索,满足阈值条件的则作为查询特征的匹配特征,保存在 FWD_MATCH 属性中.算法 1 详细描述了上述算法的设计过程.

算法 1. 基于 ACIO 和 HKD-tree 的特征匹配算法.

Input: SPI, features detected in DoG space and located precisely by interpolation with calculated main orientation

Output: Features with their correspondences stored in FWD_MATCH of feature structure

Step 1: Creating hierarchical structure
  partition SPI into H stripes in terms of formulas (5) and (6)
  establish stripe map bwtween reference stripes and target stripes:$map_{rk\to tk}=(s_{rk},s_{tk})(k=1,\cdots,H)$
  for K=1:H do
    build s_{tk}'s K-d tree $kdtree_k$ as usual
    store its root in *root_list*
    establish tree map between s_{tk} and $kdtree_k$:$map_{tk\to kdtree(k)}=(s_{tk},root_list)$
  end for
Step 2: Matching
  for K=1:H do
    for i=1:n do //n is the feature number of s_{rk}
      get feature r_i's target stripe s_{tk} by stripe map $map_{rk\to tk}$
      get s_{tk}'s K-d tree $kdtree_k$ by tree map $map_{tk\to kdtree(k)}$
      search 1-nearest neighbour and 2-nearest neighbour
      if found and ratio of KNN less than threshold and satisfying ACIO, then
        r_i's FWD_MATCH set to 1-nearest neighbor
      end if
    end for
  end for
end algorithm

3 实验与分析

为了验证本文所提方法的有效性,接下来将本文方法分别与 Brute Force 方法、SKD-Tree 方法及 CasCash 方法进行对比.Brute Force 方法是进行穷尽式搜索(exhaustive search),在速度上较慢,但其准确性比较高,因此可以作为比较的基准.SKD-Tree 方法是 OpenCV 中所使用的方法,完整地实现了文献[4]中的算法并进行了优化,具有一定的代表性.CasCash 方法是最新的方法,因此,本文将与以上 3 种方法在准确性和快速性上进行比较.

© 中国科学院软件研究所 http://www.jos.org.cn

3.1 实验环境

实验中,4 种方法用 C 语言实现,运行于 DELL PRECISION T3500 PC 机上,操作系统是 Win7 SP1,处理器是 Intel Xeon,W3503@ 2.40HZ 2.40HZ,内存为 4GB;采用的图像数据是由 Middlebury 提供的标准数据库中的立体图像对(如图 4 所示),6 种数据库的图像尺寸为 1390×1110(前 4 种)和 1342×1110(后两种).

Fig.4 Six databases from middlebury: Art, book, doll, moebius, laundry, reindeer

图 4 Middlebury 的 6 种数据库:艺术、书、玩偶、莫比斯、洗衣房、驯鹿

3.2 结果分析

3.2.1 速度分析

实验中,特征点的提取采用 DoG 方法,且 4 种匹配方法都处理同样的特征点集.为了保证实验结果更加公正,在每一个数据库上,每一种匹配方法都运行 10 次,然后取结果的平均值.表 1 是 4 种匹配方法的运行时间.图 5 是表 1 中运行时间的直观显示.

Table 1 Matching time of four methods running on different databases (Unit: ms)

表 1 4 种匹配方法在不同数据上的运行时间 (单位:毫秒)

方法	数据库名称(特征点数)					
	Art (4647)	Book (5268)	Doll (6780)	Laundry (9297)	Moebius (4890)	Reindeer (5048)
Brute Force	3 806.50	5 476.60	8 637.25	17 043.67	4 668.06	4 024.47
SKD-Tree	411.52	537.79	691.05	992.23	477.16	380.59
CasHash	57.68	76.91	112.54	633.77	67.57	103.55
HKD-Tree	13.07	52.85	44.26	113.06	23.5	12.70

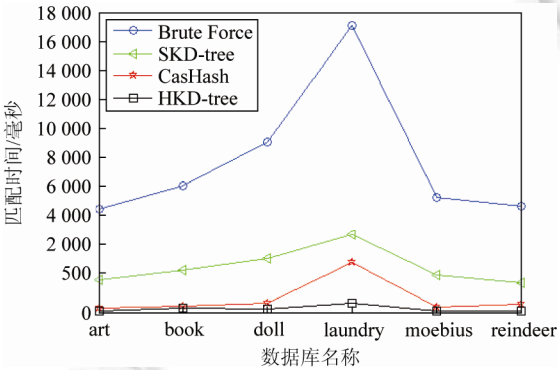

Fig.5 Matching time of four methods running on different databases

图 5 4 种匹配方法在不同数据库上的运行时间

从表中可以看出,Brute Force 方法匹配时间最长,SKD-tree 方法较之快了约一个量级,而最新方法 CasHash 的确比标准方法要快,但是随着特征点数目的增多,这种优势在逐渐下降,如在处理数据库 Laundry 时可以看出

© 中国科学院软件研究所 http://www.jos.org.cn

这一点.本文方法在所有数据库上都获得了很好的表现,同样,特征点增多时处理时间有所增加,但是仍然比标准方法约快一个量级,同时比最新方法快数倍.

3.2.2 精度分析

精度测量通常采用的方法有 ROC(receiver operating characteristics)和 Recall-Precision 两种[20],但 ROC 曲线更偏向于分类的任务,所以对特征点匹配精度的衡量采用后者比较合适.在 Recall-Precision 中,主要分析 recall 和 1-precision 两种数值在不同阈值下的分布,其定义为

$$recall = \frac{number\ of\ correct_positives}{total\ number\ of\ positives} \tag{7}$$

$$1-precision = \frac{number\ of\ false_positives}{total\ number\ of\ matches(\text{correct or false})} \tag{8}$$

从公式(7)中可以看出,recall 越大,表明正确匹配的特征点数目越多,而公式(8)表明,1-precision 越小,错误匹配的特征点相对来说也越少,总体上 recall vs. 1-precision 曲线越靠近坐标系的左上角表明性能是最好的(但是 recall 和 1-precision 之间并没有直接的关系[21],因为它们参考的对象不一样).

图 6 是 4 种方法在不同数据库上的匹配结果.图 6 中,蓝色圆圈曲线是 Brute Force 方法的测量结果,在每一种数据库上它都有稳定的表现,文献[1]中也证实了这一点,同时级联哈希方法略微优于标准方法.

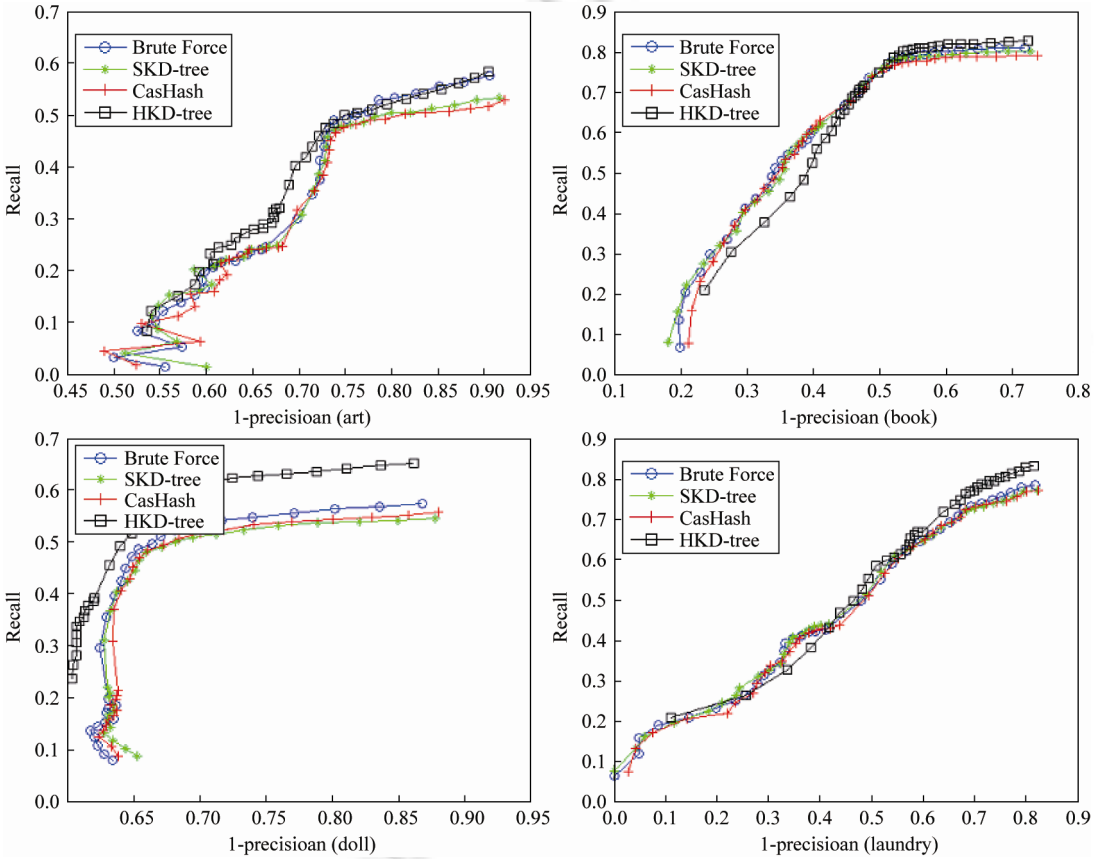

Fig.6 Matching accuracy of four methods running on different databases

图 6 4 种匹配方法在不同数据库上的匹配精度

© 中国科学院软件研究所 http://www.jos.org.cn

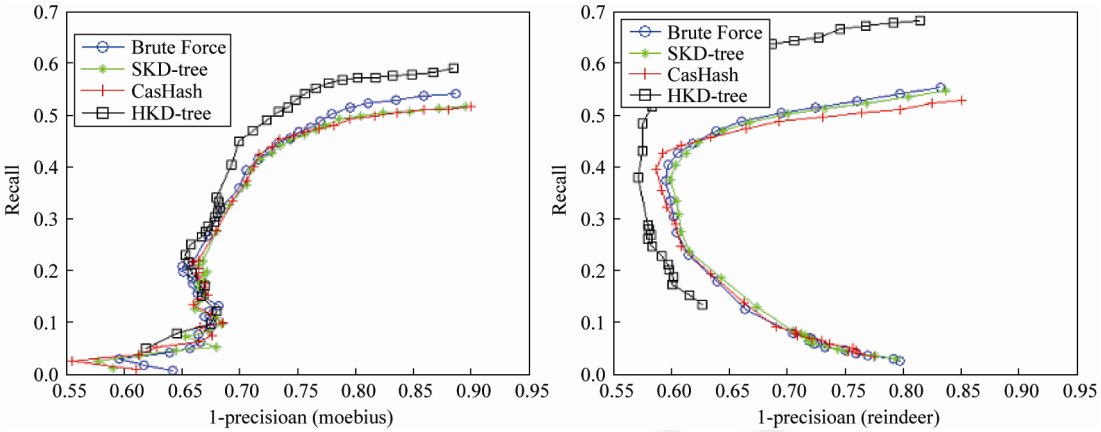

Fig.6 Matching accuracy of four methods running on different databases (Continued)

图 6 4 种匹配方法在不同数据库上的匹配精度(续)

本文方法在 art、book、laundry 这 3 种数据库上和 Brute Force 方法相差无几,而在剩余几个数据库上的表现更出色一些,原因是在计算特征点相似度时采用 128 维特征向量的“距离”,此时满足阈值的一对特征点作为一个匹配,但是它们的空间位置如果不是物理意义上的同一点,那么这对特征点则属于错误的匹配.由公式(8)的定义可知,当错误匹配增多时,1-precision 会变大,导致曲线向右侧偏移,同时正确的匹配相对来说也会减少,由公式(7)可以得出,recall 的值会变小,从而使得曲线靠近 X 轴,因此,曲线整体会处于下方.Brute Force 方法虽然完成了全部的比对,但是由于没有考虑空间位置关系,从而导致出错的几率有所上升.其他两种方法也存在同样的问题,而本文正是基于这样的出发点提出了方向大约一致(ACIO)约束关系,将不属于相近物理区域的特征点分开(即层次结构的意义所在),达到避免在向量距离的计算上属于匹配的,而在真正的物理位置上却不是同一点的误判现象发生.这样既降低了误匹配率,也减少了匹配的次数,而这也正是本文方法能够提高匹配速度和精度的主要原因.

4 结 论

本文主要提出了方向大约一致(ACIO)约束关系并给出了完整的定义,然后通过对标准 K-d 树结构进行分析,提出了具有层次结构的 K-d 树(HKD-tree),它能够有效地减少误匹配,从而降低匹配时间和提高匹配精度,同时提出基于 ACIO 和 HKD-Tree 的匹配算法.最后的实验结果表明,本文的方法在匹配速率上优于最新的级联哈希方法,比标准方法快了一个数量级,同时在匹配精度上有所提升.

从实验结果可以看出,本文方法在匹配精度上没有明显优势,这有待于进一步的研究和提高.同时,在建立层次结构时,如果存在透视变换,本文方法不能直接使用,这是我们后续要研究的重点.

References:

[1] Cheng J, Leng C, Wu JX, Cui HN, Lu HQ. Fast and accurate image matching with cascade hashing for 3d reconstruction. In: Proc. of the 2014 IEEE Conf. on Computer Vision and Pattern Recognition. Washington: IEEE, 2014. 1–8. [doi: 10.1109/CVPR.2014.8]

[2] Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proc. of the 2005 IEEE Conf. on Computer Vision and Pattern Recognition. Washington: IEEE, 2005. 886–893. [doi: 10.1109/CVPR.2005.177]

[3] Lowe DG. Object recognition from local scale-invariant features. In: Proc. of the 7th IEEE Int'l Conf. on Computer Vision. Washington: IEEE, 1999,2:1150–1157. [doi: 10.1109/ICCV.1999.790410]

[4] Lowe DG. Distinctive image features from scale-invariant keypoints. Int'l Journal of Computer Vision, 2004,60(2):91–110. [doi: 10.1023/B:VISI.0000029664.99615.94]

[5] Snavely N, Seitz SM, Szeliski R. Photo tourism: Exploring photo collections in 3D. ACM Trans. on Graphics (TOG), 2006,25(3): 835–846. [doi: 10.1145/1141911.1141964]

© 中国科学院软件研究所 http://www.jos.org.cn

[6] Philbin J, Chum O, Isard M, Sivic J, Zisserman A. Object retrieval with large vocabularies and fast spatial matching. In: Proc. of the 2007 IEEE Conf on Computer Vision and Pattern Recognition. Washington: IEEE, 2007. 1–8. [doi: 10.1109/CVPR.2007.383172]

[7] Mikolajczyk K, Leibe B, Schiele B. Multiple object class detection with a generative model. In: Proc. of the 2006 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. Washington: IEEE, 2006,1:26–36. [doi: 10.1109/CVPR.2006.202]

[8] Ferrari V, Tuytelaars T, Van Gool L. Simultaneous object recognition and segmentation by image exploration. In: Pajdla T, ed. Proc. of the Computer Vision-ECCV 2004. Heidelberg: Springer-Verlag, 2004. 40–54. [doi: 10.1007/978-3-540-24670-1_4]

[9] Arth C, Leistner C, Bischof H. Robust local features and their application in self-calibration and object recognition on embedded systems. In: Proc. of the 2007 IEEE Conf. on Computer Vision and Pattern Recognition. Washington: IEEE, 2007. 1–8. [doi: 10.1109/CVPR.2007.383419]

[10] Li FF, Perona P. A Bayesian hierarchical model for learning natural scene categories. In: Proc. of the 2005 IEEE Conf. on Computer Vision and Pattern Recognition. Washington: IEEE, 2005. 524–531. [doi: 10.1109/CVPR.2005.16]

[11] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: Proc. of the 2006 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. Washington: IEEE, 2006,2: 2169–2178. [doi: 10.1109/CVPR.2006.68]

[12] Bay H, Tuytelaars T, Van Gool L. Surf: Speeded up robust features. In: Leonardis A, ed. Proc. of the Computer vision–ECCV 2006. Heidelberg: Springer-Verlag, 2006. 404–417. [doi: 10.1007/11744023_32]

[13] Toews M, Wells W. SIFT-Rank: Ordinal description for invariant feature correspondence. In: Proc. of the 2009 IEEE Conf on Computer Vision and Pattern Recognition. Washington: IEEE, 2009. 172–177. [doi: 10.1109/CVPR.2009.5206849]

[14] Calonder M, Lepetit V, Strecha C, Fua P. Brief: Binary robust independent elementary features. In: Daniilidis K, ed. Proc. of the Computer Vision–ECCV 2010. Heidelberg: Springer-Verlag, 2010. 778–792. [doi: 10.1007/978-3-642-15561-1_56]

[15] Rublee E, Rabaud V, Konolige K, Bradski G. ORB: An efficient alternative to SIFT or SURF. In: Proc. of the 2011 IEEE Int'l Conf. on Computer Vision. Washington: IEEE, 2011. 2564–2571. [doi: 10.1109/ICCV.2011.6126544]

[16] Li B, Liu L, Wei ZQ. A strong robust real-time image matching algorithm. Ruan Jian Xue Bao/Journal of Software, 2014,25(7): 1583–1592 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/4450.htm. [doi: 10.13328/j.cnki.jos.004450]

[17] Pele O, Werman M. A linear time histogram metric for improved shift matching. In: Forsyth D, ed. Proc. of the Computer Vision–ECCV 2008. Heidelberg: Springer-Verlag, 2008. 495–508. [doi: 10.1007/978-3-540-88690-7_37]

[18] Ling HB, Okada K. Diffusion distance for histogram comparison. In: Proc. of the 2006 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. Washington: IEEE, 2006,1:246–253. [doi: 10.1109/CVPR.2006.99]

[19] Werman M, Peleg S, Melter R, Kong TY. Bipartite graph matching for points on a line or a circle. Journal of Algorithms, 1986,7(2): 277–284. [doi: 10.1016/0196-6774(86)90009-X]

[20] Ke Y, Sukthankar Y. PCA-SIFT: A more distinctive representation for local image descriptors. In: Proc. of the 2004 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. Washington: IEEE, 2004,2:II-506–II-513. [doi: 10.1109/CVPR.2004.1315206]

[21] Mikolajczyk K, Schmid C. A performance evaluation of local descriptors. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2005,27(10):1615–1630. [doi: 10.1109/TPAMI.2005.188]

[22] Gilinsky A, Zelnik-Manor L. SIFTpack: A compact representation for efficient SIFT matching. In: Proc. of the 2013 IEEE Int'l Conf. on Computer Vision. Washington: IEEE, 2013. 777–784. [doi: 10.1109/ICCV.2013.101]

[23] Strecha C, Bronstein AM, Bronstein MM, Fua P. LDAHash: Improved matching with smaller descriptors. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2012,34(1):66–78. [doi: 10.1109/TPAMI.2011.103]

[24] Liu C, Yuen J, Torralba A, Sivic J, Freeman WT. Sift flow: Dense correspondence across different scenes. In: Forsyth D, Torr P, Zisserman A, eds. Proc. of the Computer Vision–ECCV 2008. Heidelberg: Springer-Verlag, 2008. 28–42. [doi: 10.1007/978-3-540-88690-7_3]

[25] Ghamdi MA, Zhang L, Gotoh Y. Spatio-Temporal SIFT and its application to human action classification. In: Fusiello A, ed. Proc. of the Computer Vision–ECCV 2012, Workshops and Demonstrations. Heidelberg: Springer-Verlag, 2012. 301–310. [doi: 10.1007/978-3-642-33863-2_30]

[26] Lopes APB, Oliveira RS, de Almeida JM, de Arnaldo AA. Spatio-Temporal frames in a bag-of-visual-features approach for human actions recognition. In: Proc. of the 2009 XXII Brazilian Symp. on Computer Graphics and Image Processing. Washington: IEEE, 2009. 315–321. [doi: 10.1109/SIBGRAPI.2009.17]

[27] Kanade T, Okutomi M. A stereo matching algorithm with an adaptive window: Theory and experiment. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1994,16(9):920–932. [doi: 10.1109/34.310690]

© 中国科学院软件研究所 http://www.jos.org.cn

[28] Veksler O. Stereo matching by compact windows via minimum ratio cycle. In: Proc. of the 8th IEEE Int'l Conf. on Computer Vision, 2001. Washington: IEEE, 2001,1:540–547. [doi: 10.1109/ICCV.2001.937563]

[29] Veksler O. Fast variable window for stereo correspondence using integral images. In: Proc. of 2003 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. Washington: IEEE, 2003,1:I-556-I-561. [doi: 10.1109/CVPR.2003.1211403]

[30] Wang K. Adaptive stereo matching algorithm based on edge detection. In: Proc. of the 2004 Int'l Conf. on Image Processing. Washington: IEEE, 2004,2:1345–1348. [doi: 10.1109/ICIP.2004.1419748]

[31] Yoon KJ, Kweon IS. Adaptive support-weight approach for correspondence search. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2006,28(4):650–656. [doi: 10.1109/TPAMI.2006.70]

[32] Banks J, Bennamoun M, Corke P. Non-Parametric techniques for fast and robust stereo matching. In: Proc. of the IEEE Region 10 Annual Conf. on Speech and Image Technologies for Computing and Telecommunications. Washington: IEEE, 1997,1:365–368. [doi: 10.1109/TENCON.1997.647332]

[33] Zabih R, Woodfill J. Non-Parametric local transforms for computing visual correspondence. In: Eklundh JO, ed. Proc.of the Computer Vision—ECCV'94. Heidelberg: Springer-Verlag, 1994,2:151–158. [doi: 10.1007/BFb0028345]

[34] Cai J. Integration of optical flow and dynamic programming for stereo matching. IET Image Processing, 2012,6(3):205–212. [doi: 10.1049/iet-ipr.2010.0070]

[35] Nguyen M, Chan YH, Delmas P, Gimel'farb G. Symmetric dynamic programming stereo using block matching guidance. In: Proc. of the 28th Int'l Conf. of Image and Vision Computing New Zealand (IVCNZ). Washington: IEEE, 2013. 88–93. [doi: 10.1109/IVCNZ.2013.6726997]

[36] Taniai T, Matsushita Y, Naemura T. Graph cut based continuous stereo matching using locally shared labels. In: Proc. of the 2014 IEEE Conf. on Computer Vision and Pattern Recognition. Washington: IEEE, 2014. 1613–1620. [doi: 10.1109/CVPR.2014.209]

[37] Zhang H, Cheng FY, Yuan D, Li YC, Sun MG. Stereo matching with global edge constraint and graph cuts. In: Proc. of the 21st Int'l Conf. on Pattern Recognition (ICPR). Washington: IEEE, 2012. 372–375.

[38] Sun J, Zheng NN, Shum HY. Stereo matching using belief propagation. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2003,25(7):787–800. [doi: 10.1109/TPAMI.2003.1206509]

[39] Ahmadzadeh A, Madani H, Jafari K, Jazi FS, Daneshpajouh S, Gorgin S. Fast and adaptive BP-based multi-core implementation for stereo matching. In: Proc. of the 11th IEEE/ACM Int'l Conf. on Formal Methods and Models for Codesign (MEMOCODE). Washington: IEEE, 2013. 135–138.

[40] Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int'l Journal of Computer Vision, 2002,47(1/3):7–42. [doi: 10.1023/A:1014573219977]

[41] Heiko H, Scharstein D. Evaluation of cost functions for stereo matching. In: Proc. of the 2007 IEEE Conf. on Computer Vision and Pattern Recognition. IEEE, 2007. 1–8. [doi: 10.1109/CVPR.2007.383248]

附中文参考文献:

[16] 李兵,刘磊,魏志强.一种具有强实时性、强鲁棒性的图像匹配算法.软件学报,2014,25(7):1583–1592. http://www.jos.org.cn/1000-9825/4450.htm. [doi: 10.13328/j.cnki.jos.004450]

张贵安(1985－),男,安徽当涂人,博士生,主要研究领域为计算机视觉.

童倩倩(1989－),女,博士生,主要研究领域为人机交互,机器学习.

袁志勇(1963－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为虚拟现实与人机交互,计算机视觉,机器学习,模式识别.

廖祥云(1989－),男,博士生,CCF 学生会员,主要研究领域为计算机仿真,虚拟现实,图像处理.

© 中国科学院软件研究所 http://www.jos.org.cn